

## TRACY–WIDOM LIMIT FOR KENDALL’S TAU

BY ZHIGANG BAO<sup>1</sup>

*Hong Kong University of Science and Technology*

In this paper, we study a high-dimensional random matrix model from nonparametric statistics called the Kendall rank correlation matrix, which is a natural multivariate extension of the Kendall rank correlation coefficient. We establish the Tracy–Widom law for its largest eigenvalue. It is the first Tracy–Widom law for a nonparametric random matrix model, and also the first Tracy–Widom law for a high-dimensional U-statistic.

**1. Introduction.** Let  $\mathbf{w} = (w_1, \dots, w_p)'$  be a  $p$ -dimensional random vector. We assume that all the components of  $\mathbf{w}$  are independent continuous random variables. We do not require the components to be identically distributed, and no moment assumption on the components of  $\mathbf{w}$  is needed. Let  $\mathbf{w}_j = (w_{1j}, \dots, w_{pj})'$ ,  $j \in \llbracket 1, n \rrbracket$  be  $n$  i.i.d. samples of  $\mathbf{w}$ . Hereafter, we use the notation  $\llbracket a, b \rrbracket := [a, b] \cap \mathbb{Z}$ . We also denote by  $W = (w_{ij})_{p,n}$  the data matrix. In the paper, we assume that  $p$  and  $n$  are comparable. More specifically, we assume

$$(1.1) \quad p = p(n), \quad c_n := \frac{p}{n} \rightarrow c \in (0, \infty) \quad \text{if } n \rightarrow \infty,$$

for some positive constant  $c$ .

From the data matrix  $W$ , we can further construct a matrix model called *Kendall rank correlation matrix*, originating from nonparametric statistics. The definition is detailed as follows.

**1.1. Kendall rank correlation matrix.** Recall the data matrix  $W = (w_{ij})_{p,n}$ . For any given  $k \in \llbracket 1, p \rrbracket$ , we denote

$$(1.2) \quad v_{k,(ij)} := \text{sign}(w_{ki} - w_{kj}) \quad \forall i \neq j$$

and let

$$(1.3) \quad \boldsymbol{\theta}_{(ij)} := \frac{1}{\sqrt{M}}(v_{1,(ij)}, \dots, v_{p,(ij)})',$$

where for brevity we set

$$M \equiv M(n) := \frac{n(n-1)}{2}.$$

---

Received March 2018; revised August 2018.

<sup>1</sup>Supported in part by Hong Kong RGC GRF Grant 16300618.

*MSC2010 subject classifications.* Primary 60B20, 62G10; secondary 62H10, 15B52, 62H25.

*Key words and phrases.* Tracy–Widom law, largest eigenvalue, nonparametric statistics, U-statistics, random matrices.

The Kendall rank correlation matrix is defined as the following sum of  $M$  rank-one matrices:

$$(1.4) \quad K \equiv K_n := \sum_{i < j} \boldsymbol{\theta}_{(ij)} \boldsymbol{\theta}'_{(ij)} = \Theta \Theta'.$$

Here we denote by

$$(1.5) \quad \Theta := (\boldsymbol{\theta}_{(12)}, \dots, \boldsymbol{\theta}_{(1n)}, \boldsymbol{\theta}_{(23)}, \dots, \boldsymbol{\theta}_{(2n)}, \dots, \boldsymbol{\theta}_{(n-1,n)}).$$

Observe that the rank-one matrices  $\boldsymbol{\theta}_{(ij)} \boldsymbol{\theta}'_{(ij)}$ 's are not independent. For instance,  $\boldsymbol{\theta}_{(ij)} \boldsymbol{\theta}'_{(ij)}$  and  $\boldsymbol{\theta}_{(ik)} \boldsymbol{\theta}'_{(ik)}$  are correlated even if  $j \neq k$ . Moreover,  $K$  is a  $p \times p$  matrix, and its  $(a, b)$ -entry is

$$K_{ab} = \frac{1}{M} \sum_{i < j} v_{a,(ij)} v_{b,(ij)} = \frac{1}{M} \sum_{i < j} \text{sign}(w_{ai} - w_{aj}) \text{sign}(w_{bi} - w_{bj}),$$

which is exactly the Kendall rank correlation coefficient between the samples of  $w_a$  and those of  $w_b$ . Hence, the matrix  $K$  is a natural multivariate extension of the Kendall rank correlation coefficient.

1.2. *Motivation.* Since the seminal work of Marchenko and Pastur [28], the spectral properties of large dimensional sample covariance matrix and its variants have attracted enormous attention. In [28], the famous Marchenko–Pastur law (MP-law) for the global spectral distribution of the sample covariance matrices has been raised. On the local scale, Johnstone [21] proved the Tracy–Widom law (TW law) for the largest eigenvalue of the real Gaussian sample covariance matrix (Wishart matrix) in the null case, that is, the population covariance matrix is  $I_p$ . Since the largest eigenvalue plays a fundamental role in principal component analysis (PCA), the TW law can be applied to many PCA-related problems in high-dimensional scenarios. The TW law was then shown to be universal for sample covariance matrices in the null case, even under more general distribution assumptions; see [31, 32]. In [5, 30], it was also shown that the TW law holds for the (Pearson) sample correlation matrix in the null case. We also mention [12, 20, 29] as they give related results for complex sample covariance matrices. Recently, the universality was further established for more general population; see [6, 16, 23, 25].

Both the sample covariance matrix and (Pearson) sample correlation matrix are parametric models. Many spectral statistics such as the largest eigenvalue of the sample covariance matrix or correlation matrix are used for testing the hypothesis of *independence* among the entries of a random vector. The strategy is certainly feasible for Gaussian vectors. However, for non-Gaussian vectors, even in the classical large  $n$  and fixed  $p$  case, the idea of comparing population covariance matrix with diagonal matrix cannot be used for an independence test involving uncorrelated but dependent variables. On the other hand, although the TW law was shown

to be universal for sample covariance matrices, assumptions on the distribution of the matrix entries are still required to a certain extent; see for instance, the minimal moment condition in [10]. This moment requirement certainly excludes all heavy-tailed data sets. For the above reasons, a more robust nonparametric approach is needed.

In classical nonparametric statistics, the most famous statistics concerning the statistical dependence between two random variables are the Spearman rank correlation coefficient and the Kendall rank correlation coefficient, also known as Spearman’s  $\rho$  and Kendall’s  $\tau$ . Both of them have natural multivariate extensions, which are called Spearman rank correlation matrix and Kendall rank correlation matrix (cf. (1.4)), respectively. Since these models are nonparametric, all the hypothesis tests based on statistics of these models are distribution free. However, in contrast to the parametric models, the study on the spectral properties of the high-dimensional nonparametric matrices is much less. Under the null hypothesis, that is, the components of  $w$  are independent, the global spectral distributions for the Spearman rank correlation matrix and Kendall rank correlation matrix have been derived in [1] and [3], respectively. A CLT for the linear eigenvalue statistics of the Spearman rank correlation matrix has been considered in [4]. However, so far, there is no result on the local eigenvalue statistics such as the largest eigenvalue of these two nonparametric models. In this work, our aim is to establish the TW law for the Kendall rank correlation matrix. In a companion paper [8], we show that the TW law also holds for the Spearman rank correlation matrix.

Moreover, it is also well known that Kendall’s tau is a U-statistic. The spectral theory on general high-dimensional U-statistics is still unexplored, except for the global law of Kendall’s tau in [3]. The result in this paper can also be regarded as the first TW law established for a high-dimensional U-statistic. Furthermore, we expect that the method developed in this paper will, to a certain extent, have potential applications to other high-dimensional U-statistics.

1.3. *Global behavior of the spectrum.* In this subsection, we first review the result on the global law from [3]. Let  $\lambda_1(K) \geq \dots \geq \lambda_p(K)$  be  $p$  ordered eigenvalues of  $K$ . Denote the empirical spectral distribution (ESD) of  $K$  by

$$F_n^K := \frac{1}{p} \sum_{i=1}^p \delta_{\lambda_i(K)}.$$

In [3], it is proved the  $F_n^K$  is asymptotically given by a scaled and shifted MP law. To state the result in [3], we first introduce the Marchenko–Pastur law  $F_c$  (with parameter  $c$ ), whose density function is given by

$$\rho_c(x) = \frac{1}{2\pi c} \frac{\sqrt{(d_{+,c} - x)(x - d_{-,c})}}{x} \mathbb{1}(d_{-,c} \leq x \leq d_{+,c}),$$

where  $d_{\pm,c} = (1 \pm \sqrt{c})^2$ . In case  $c > 1$ , in addition,  $F_c$  has a singular part: a point mass  $(1 - c^{-1})\delta_0$ .

**THEOREM 1.1** (Theorem 1 of [3]). *Under the assumption (1.1), we have that  $F_n^K$  converges weakly (in probability) to  $F_c^K$  whose density is given by*

$$\rho_c^K(x) = \frac{3}{2} \rho_c \left( \frac{3}{2}x - \frac{1}{2} \right).$$

Hence,  $F_c^K(x) = F_c(\frac{3}{2}x - \frac{1}{2})$ .

Further, replacing  $c$  by  $c_n$ , we denote by  $\rho_{c_n}, \rho_{c_n}^K, F_{c_n}, F_{c_n}^K, d_{\pm, c_n}$  the analogues of  $\rho_c, \rho_c^K, F_c, F_c^K, d_{\pm, c}$ , respectively. Further, we introduce the shorthand notation

$$(1.6) \quad \lambda_{\pm, c_n} := \frac{2}{3}d_{\pm, c_n} + \frac{1}{3}.$$

**1.4. Main results.** To state our main results, we denote by  $Q := \frac{1}{n} \mathcal{X} \mathcal{X}'$  a Wishart matrix, where  $\mathcal{X}$  is a  $p \times n$  data matrix with i.i.d.  $N(0, 1)$  variables. Let  $\lambda_i(Q)$  be the  $i$ th largest eigenvalue of  $Q$ . Our main results are as follows.

**THEOREM 1.2** (Edge universality of Kendall rank correlation matrix). *Suppose that the assumption (1.1) holds. There exist positive constants  $\varepsilon$  and  $\delta$  such that for any  $s \in \mathbb{R}$ , the following holds for all sufficiently large  $n$ :*

$$(1.7) \quad \begin{aligned} & \mathbb{P} \left( \frac{3}{2}n^{\frac{2}{3}}(\lambda_1(K) - \lambda_{+, c_n}) \leq s - n^{-\varepsilon} \right) - n^{-\delta} \\ & \leq \mathbb{P} \left( n^{\frac{2}{3}}(\lambda_1(Q) - d_{+, c_n}) \leq s \right) \\ & \leq \mathbb{P} \left( \frac{3}{2}n^{\frac{2}{3}}(\lambda_1(K) - \lambda_{+, c_n}) \leq s + n^{-\varepsilon} \right) + n^{-\delta}. \end{aligned}$$

**REMARK 1.3.** The above theorem can be extended to the joint distribution for the first  $k$  leading eigenvalues. We refer to Remark 1.4 of [31] for a similar extension for the sample covariance matrix. The extension here can be done in the same way.

From Theorem 1.2, we can get the following corollary.

**COROLLARY 1.4** (Tracy–Widom law for  $\lambda_1(K)$ ). *Under the assumption of Theorem 1.2, we have*

$$\frac{3}{2}n^{\frac{2}{3}}c_n^{\frac{1}{6}}d_{+, c_n}^{-\frac{2}{3}}(\lambda_1(K) - \lambda_{+, c_n}) \implies \text{TW}_1,$$

where  $\text{TW}_1$  stands for the Tracy–Widom law of type I.

1.5. *Proof strategy.* In the sequel, we summarize our proof strategy with a highlight on the novelties. Our proof strategy traces back to the seminal works of Erdős, Yau and Yin [14, 15], where a general framework to prove the universality of local eigenvalue statistics has been raised. Roughly speaking, the strategy in [15] for proving the edge universality consists of two major steps. First, one needs to prove a local law for the spectral distribution, from which one can get a control on the location of the eigenvalues on an optimal local scale. Second, with the aid of the local law, one needs to perform a Green function comparison between the matrix of interest and a certain reference matrix ensemble, whose edge spectral behavior is already known. In the Green function comparison step, one translates the comparison between the distributions of the largest eigenvalues of two random matrices to a comparison of their Green functions. The Green function turns out to be a more convenient object to look into, due to the simple resolvent expansion mechanism. An adaptation of this general strategy was used by Pillai and Yin in [31] to show both the bulk and edge universality of the sample covariance matrices. Especially, in [31], an extended criterion of the local law for covariance type of matrices with independent columns (or rows) was given; see Theorem 3.6 of [31]. It allows one to relax the independence assumption on the entries within each single column (or row) to a certain extent, as long as some large deviation estimates hold for certain linear and quadratic forms of each column (or row) of the data matrix; see Lemma 3.4 of [31]. This general criterion was then used in [30] and [5] to establish the edge universality of the sample correlation matrices.

In order to illustrate the new ingredients in applying the above general strategy to our model, we first introduce some notation. For any parameter  $z \in \mathbb{C}^+$ , we denote by  $G(z) = (G_{k\ell}(z)) := (K - z)^{-1}$  the Green function of  $K$  and by  $m(z) := \frac{1}{p} \text{Tr} G(z)$  the normalized trace of the Green function, which is also the Stieltjes transform of the ESD  $F_n^K$ . Let  $\underline{m}(z)$  be the Stieltjes transform of  $F_c^K$ . For our matrix  $K$ , in the step of local law, one needs to establish the following estimates:

$$(1.8) \quad |G_{k\ell}(z) - \delta_{k\ell} \underline{m}(z)| \prec \Psi(z),$$

$$(1.9) \quad |m(z) - \underline{m}(z)| \prec \frac{1}{n \text{Im} z}$$

in the domain  $\mathcal{D}(\epsilon)$  (cf. (4.3)). We also refer to (4.4) and Definition 1.5 for the definition of  $\Psi(z)$  and the notation  $\prec$ , respectively. It is now well understood that a large deviation estimate of  $\lambda_i(K)$  around its classical location can be derived from the local law. However, the large deviation estimate does not tell the TW law of  $\lambda_1(K)$  directly, although together with (1.8) and (1.9) it will serve as an important input for the proof of the TW law. As we mentioned above, for TW law, as the next step, we need to conduct a Green function comparison. In this step, we will compare the distribution function of  $\lambda_1(K)$  with that of  $\lambda_1(\tilde{K})$ , where  $\tilde{K}$  (cf. (6.1)) is a shifted covariance matrix and the law of  $\lambda_1(\tilde{K})$  is known to be  $\text{TW}_1$ .

The comparison of the distributions can be translated into the comparison of the Green functions, and it suffices to show

$$(1.10) \quad \left| \mathbb{E}F\left(n \int_{E_1}^{E_2} \operatorname{Im} m(x + \lambda_{+,c_n} + i\eta) dx\right) - \mathbb{E}F\left(n \int_{E_1}^{E_2} \operatorname{Im} \tilde{m}(x + \lambda_{+,c_n} + i\eta) dx\right) \right| \leq n^{-\delta},$$

where  $F$  is a smooth test function and  $\tilde{m}$  stands for the Stieltjes transform of the ESD of  $\tilde{K}$ . We refer to Proposition 5.1 for the setting of  $\eta$ ,  $E_1$  and  $E_2$ . The proof of (1.10) will heavily rely on (1.8) and (1.9).

As we mentioned above, the Kendall rank correlation matrix is a multivariate U-statistic. Its structure is significantly different from the sample covariance matrix or correlation matrix. Although the rows of  $\Theta$  are mutually independent, there is a strong dependence structure among the entries within each row. Consequently, both the proofs of the two steps, that is, local law and Green function comparison, require novel ideas.

The starting point of the whole proof is (a variant of) Hoeffding decomposition [18], which is already used for the global law in [3]. Specifically, for Kendall rank correlation, we can decompose  $v_{k,(ij)}$  (cf. (1.2)) as

$$(1.11) \quad v_{k,(ij)} = u_{k,(ij)} + \bar{v}_{k,(ij)},$$

where

$$(1.12) \quad u_{k,(ij)} := \mathbb{E}(\operatorname{sign}(w_{ki} - w_{kj})|w_{ki}) + \mathbb{E}(\operatorname{sign}(w_{ki} - w_{kj})|w_{kj}),$$

and we take the above as the definition of  $\bar{v}_{k,(ij)}$ . It is easy to check that  $u_{k,(ij)}$  and  $\bar{v}_{k,(ij)}$  are uncorrelated. Correspondingly, we set the  $p \times M$  matrices  $U = \frac{1}{\sqrt{M}}(u_{k,(ij)})_{k,(ij)}$  and  $\bar{V} = \frac{1}{\sqrt{M}}(\bar{v}_{k,(ij)})_{k,(ij)}$ . Hence, we have the decomposition  $\Theta = U + \bar{V}$ . In the sequel, we will call  $U$  the *linear* part of  $\Theta$ , and  $\bar{V}$  the *nonlinear* part of  $\Theta$ . It will be seen that  $UU'$  is indeed a covariance type of matrix and its spectral property can be obtained from the results on sample covariance matrices easily. However, in  $K = \Theta\Theta' = (U + \bar{V})(U + \bar{V})'$ , we also have the crossing parts  $\bar{V}U'$ ,  $U\bar{V}'$  and the purely nonlinear part  $\bar{V}\bar{V}'$ . The nonlinear term  $\bar{V}$  couples the columns of  $\Theta$  together, and makes the structure of  $K$  different from the covariance matrix.

For the step of local law, recall our tasks (1.8) and (1.9). We take the estimate of the diagonal entries  $G_{kk}$ 's as an example. By Schur complement, one can write  $G_{kk}$  in terms of a quadratic form  $\mathbf{v}_k B^{(k)} \mathbf{v}_k'$ ; see (S.38) for more details. Here  $\mathbf{v}_k$  is the  $k$ th row of  $\Theta$  and it is independent of  $B^{(k)}$ . Hence, an estimate of  $G_{kk}$  essentially boils down to a large deviation estimate of the quadratic form of  $\mathbf{v}_k$ . It turns out that although a direct large deviation estimate is enough for (1.8), it is not sufficient for later use in the Green function comparison. With Hoeffding

decomposition, we can write  $\mathbf{v}_k B^{(k)} \mathbf{v}'_k$  as a linear combination of the linear part  $\mathbf{u}_k B^{(k)} \mathbf{u}'_k$ , crossing part  $\mathbf{u}_k B^{(k)} \bar{\mathbf{v}}'_k$  and the nonlinear part  $\bar{\mathbf{v}}_k B^{(k)} \bar{\mathbf{v}}'_k$ , where  $\mathbf{u}_k$  and  $\bar{\mathbf{v}}_k$  are the  $k$ th rows of  $U$  and  $\bar{V}$ , respectively. We establish the large deviation estimates for three parts separately; see Propositions 3.1 and 3.2. It turns out that the large deviations of the last two parts are much sharper than the first part, although the sharpness for the crossing part can be seen only a posteriori. The sharper large deviation estimates for the crossing part and nonlinear part will be crucial in Green function comparison. The proof of Proposition 3.2 will be the major task in this step. The matrices  $U$  and  $\bar{V}$  are only uncorrelated rather than independent, and so are the entries within  $\bar{V}$ . To prove Proposition 3.2, we need to perform a martingale concentration argument. With these large deviation estimates, we then prove the local law, by pursuing the strategy in [15] and [31].

For Green function comparison (1.10), we further decompose it into two steps. We call the first step as *decoupling*, and the second step as *first-order approximation*. In the *decoupling* step, we compare  $K = (U + \bar{V})(U + \bar{V})'$  with  $\hat{K} = (U + H)(U + H)'$ , where  $H = (h_{k,(ij)})$  is a  $p \times M$  Gaussian matrix with i.i.d.  $h_{k,(ij)} \sim N(0, \frac{1}{3M})$  and it is independent of  $U$ . This step allows us to decouple the dependent (although uncorrelated) pair  $(U, \bar{V})$  by studying the independent pair  $(U, H)$  instead. For the Green function comparison between  $K$  and  $\hat{K}$ , we use a swapping strategy via replacing one row of  $\bar{V}$  by that of  $H$  at each time and compare the Green functions step by step. Such a replacement strategy has been previously used in [31], and also [5, 6, 30]. However, such a comparison involves high order moments of the quadratic forms of  $\mathbf{v}_k$  and  $\hat{\mathbf{v}}_k$ , where  $\hat{\mathbf{v}}_k$  represents the  $k$ th row of  $U + H$ . Roughly speaking, the comparison requires the first three moments of  $\mathbf{v}_k B \mathbf{v}'_k$  and  $\hat{\mathbf{v}}_k B \hat{\mathbf{v}}'_k$  and their variants to match, up to sufficiently small errors. Here  $B$  is certain matrix independent of both  $\mathbf{v}_k$  and  $\hat{\mathbf{v}}_k$ . Although the entries in  $\bar{V}$  and those in  $H$  have the same covariance structure, their higher order moments do not match. In addition, although the entries in  $U$  and those in  $\bar{V}$  are uncorrelated, they are dependent at high orders. One key point in the comparison of the moments of  $\mathbf{v}_k B \mathbf{v}'_k$  and those of  $\hat{\mathbf{v}}_k B \hat{\mathbf{v}}'_k$  is to show that the high order correlation between the entries in  $U$  and  $\bar{V}$  is negligible. This fact heavily relies on the sharper large deviations for the crossing part and nonlinear part in Proposition 3.2. In the *first-order approximation* step, we further compare  $\hat{K} = (U + H)(U + H)'$  with the random matrix  $\tilde{K}$ . In this step, we approximate all the terms with the matrix  $H$  involved by the deterministic  $\frac{1}{3}I_p$ . The Green function comparison between  $\hat{K}$  and  $\tilde{K}$  will be done with a continuous interpolation between two matrices. Similar idea of continuous interpolation was previously used for the Green function comparison in [24, 25].

1.6. *Notation and organization.* We first need the following definition from [13].

DEFINITION 1.5. Let  $X \equiv X^{(n)}$  and  $Y \equiv Y^{(n)}$  be two sequences of nonnegative random variables. We say that  $\mathcal{Y}$  stochastically dominates  $X$  if, for all (small)  $\epsilon > 0$  and (large)  $D > 0$ ,

$$(1.13) \quad \mathbb{P}(X^{(n)} > n^\epsilon Y^{(n)}) \leq n^{-D},$$

for sufficiently large  $n \geq n_0(\epsilon, D)$ , and we write  $X \prec Y$  or  $X = O_{\prec}(Y)$ . When  $X^{(n)}$  and  $Y^{(n)}$  depend on a parameter  $v \in V$  (typically an index label or a spectral parameter), then  $X(v) \prec Y(v)$ , uniformly in  $v \in V$ , means that the threshold  $n_0(\epsilon, D)$  can be chosen independently of  $v$ . We also use the notation  $X^{(n)} \prec Y^{(n)}$  if  $X^{(n)} \leq n^\epsilon Y^{(n)}$  deterministically for any given (small)  $\epsilon > 0$ . Finally, we say that an event  $\mathcal{E} \equiv \mathcal{E}_n$  holds with high probability if: for any fixed  $D > 0$ , there exists  $n_0(D) > 0$ , such that for all  $n \geq n_0(D)$  we have

$$\mathbb{P}(\mathcal{E}) \geq 1 - n^{-D}.$$

In the case that the nonnegative random variable  $X$  satisfies the stochastic bound  $X \prec Y$  and the deterministic bound  $X \leq N^k Y$  for some nonnegative integer  $k$  and nonnegative  $Y$ , we can easily conclude that  $\mathbb{E}X^p \prec \mathbb{E}Y^p$  for any given  $p \geq 0$ . We use the symbols  $O(\cdot)$  and  $o(\cdot)$  for the standard big-O and little-o notation. We use  $C$  to denote strictly positive constant that does not depend on  $N$ . Its value may change from line to line. For any matrix  $A$ , we denote by  $\|A\|$  its operator norm, while for any vector  $\mathbf{a}$ , we use  $\|\mathbf{a}\|$  to denote its  $\ell^2$ -norm. Further, we use  $\|\mathbf{a}\|_\infty$  to represent the  $\ell^\infty$ -norm of a vector. In addition, we use double brackets to denote index sets, that is, for  $n_1, n_2 \in \mathbb{R}$ ,  $[[n_1, n_2]] := [n_1, n_2] \cap \mathbb{Z}$ . The notation  $\mathbb{1}(\cdot)$  will be used to denote the indicator function. We also use  $\mathbf{1}$  to represent the all-one vector, whose dimension may change from one to another.

The paper is organized as follows: In Section 2, we will present a simulation study to show that the testing statistic of the largest eigenvalue of the Kendall rank correlation matrix has good performance in the independence test. In Section 3, we will state some large deviation estimates which will be used in the later sections. In Section 4, we will state a local law of  $K$ . In Section 5, we will compare the Green functions of  $K$  and  $\widehat{K}$ , where the latter has independent linear and “nonlinear” parts. In Section 6, we further compare the Green functions of  $\widehat{K}$  and  $\widetilde{K}$ , where the latter is a shift of the linear part only. Section 7 will be devoted to the final proof of Theorem 1.2 and Corollary 1.4. The proofs of the large deviation bounds, the local law and some technical lemmas will be stated in the supplementary material [7]. In addition, we also present more simulation results in [7].

**2. Application and simulation study.** In this section we apply the  $TW_1$  law for  $K$  to test the complete independence of the components of the random vector  $\mathbf{w} = (w_1, \dots, w_p)^t$ . We also compare the performance of our statistic, that is,  $\lambda_1(K)$ , with some other statistics in the literature. From the  $n$  samples of  $\mathbf{w}$ , that is,  $\mathbf{w}_1, \dots, \mathbf{w}_n$ , we can define three types of correlation matrices: Pearson correlation

matrix ( $R$ ), Spearman rank correlation matrix ( $S$ ) and Kendall rank correlation matrix ( $K$ ). By definition, the matrix entries  $R_{ij}$ ,  $S_{ij}$  and  $K_{ij}$  are the Pearson, Spearman and Kendall correlation coefficient between samples of  $w_i$  and  $w_j$ , respectively. Denote by  $\lambda_1(A)$  the largest eigenvalue of  $A$ , for  $A = R, S$  and  $K$ . We will consider 7 statistics constructed from  $R, S$  and  $K$ . They are defined as follows:

- (i)  $T_1 = \frac{\text{Tr } R^2 - a_R}{b_R}$  (see [17]);
- (ii)  $T_2 = \frac{\text{Tr } S^2 - a_S}{b_S}$  (see [4]);
- (iii)  $T_3 = n \left( \max_{1 \leq i < j \leq p} |R_{ij}| \right)^2 - 4 \log n + \log \log n$  (see [19]);
- (iv)  $T_4 = n \left( \max_{1 \leq i < j \leq p} \left| \frac{p}{n} S_{ij} \right| \right)^2 - 4 \log p + \log \log p$  (see [33]);
- (v)  $T_5 = n^{\frac{2}{3}} c_n^{\frac{1}{6}} d_{+,c_n}^{-\frac{2}{3}} (\lambda_1(R) - d_{+,c_n})$  (see [5, 30]);
- (vi)  $T_6 = n^{\frac{2}{3}} c_n^{\frac{1}{6}} d_{+,c_n}^{-\frac{2}{3}} (\lambda_1(S) - d_{+,c_n})$  (see [8]);
- (vii)  $T_7 = \frac{3}{2} n^{\frac{2}{3}} c_n^{\frac{1}{6}} d_{+,c_n}^{-\frac{2}{3}} (\lambda_1(K) - \lambda_{+,c_n})$  (see Corollary 1.4),

where the parameters  $a_R, b_R, a_S$  and  $b_S$  will be explained later. We briefly describe the limiting distributions of the above statistics under the null hypothesis, that is,  $w_1, \dots, w_p$  are independent. The limiting null distributions of  $T_1$  and  $T_2$  are both  $N(0, 1)$ . The CLT for  $T_1$  is derived in [17] under a four moment assumption, and that for  $T_2$  is established in [4] for arbitrary random vector with continuous distribution. We mention that both [17] and [4] give CLT of linear eigenvalue statistics for more general test functions. Here we choose the test function  $f(x) = x^2$  for simplicity. The explicit forms of the centering constants  $a_R$  and  $a_S$  and also those for the scaling constants  $b_R$  and  $b_S$  can be found in Theorem 3.1 of [17] and Theorem 1.1 of [4]. Under a moment condition  $\mathbb{E}|w_i|^{30-\varepsilon} < \infty$  with some small constant  $\varepsilon > 0$ , the limiting null distribution of  $T_3$  is derived in [19], and it admits the following c.d.f.:  $F_{T_3}(x) = \exp(-(c^2\sqrt{8\pi})^{-1}e^{-y/2})$ . Similarly, the limiting null distribution of  $T_4$  (cf. [33]) is given by  $F_{T_4}(x) = \exp(-(8\pi)^{-1/2}e^{-y/2})$ . Since  $T_4$  is nonparametric, the above limiting law does not require moment assumption. The limiting null distributions of  $T_5, T_6, T_7$  are all given by  $TW_1$  law. In [5, 30], the  $TW_1$  law is established for  $R$ , assuming that  $w_i$ 's have sub-exponential tails. Again, since  $T_6$  and  $T_7$  are constructed from nonparametric matrices, their limiting laws do not require any moment assumption on  $w_i$ 's.

In the sequel, we denote by  $\text{Cauchy}(0, 1)$  the Cauchy distribution with location parameter 0 and scale parameter 1. We further denote by  $t(4)$  the student's  $t$ -distribution with degrees of freedom 4. We will consider three null hypotheses

with the nominal significance level  $\alpha = 5\%$ , for  $N(0, 1)$ , Cauchy(0, 1) and  $t(4)$  variables, respectively:

- $H_{0,1}$ :  $w_i$ 's are i.i.d.  $N(0, 1)$  variables;
- $H_{0,2}$ :  $w_i$ 's are i.i.d. Cauchy(0, 1) variables;
- $H_{0,3}$ :  $w_i$ 's are i.i.d.  $t(4)$  variables.

For each null hypothesis  $H_{0,i}$ ,  $i = 1, 2, 3$ , we consider two types of alternatives: (i) the alternative of one large disturbance, denoted by  $H_{a,i-1}$ ; (ii) the alternative of many small disturbances, denoted by  $H_{a,i-2}$ . Specifically, for some parameters  $\delta \in (0, 1]$  and  $\tau_1, \tau_2, \tau_3 > 0$ , we set:

- $H_{a,1-1}$ :  $\mathbf{w} \sim N_p(0, I_p + A)$ , where  $A = (a_{ij})_{p \times p}$  with  $a_{ij} = 0$  for all  $i, j$  except for  $a_{12} = a_{21} = \delta$ .
- $H_{a,1-2}$ :  $\mathbf{w} \sim N_p(0, I_p + B)$ , where  $B = (b_{ij})_{p \times p}$  with  $b_{ij} = \frac{\tau_1}{p}$  for all  $i, j$ .
- $H_{a,2-1}$ : Let  $\{x_i\}_{i=1}^p$  be i.i.d. Cauchy(0, 1). We set  $w_1 = x_1 + \delta x_2$ ,  $w_2 = \delta x_1 + x_2$  and  $w_i = x_i$  for all  $i \neq 1, 2$ .
- $H_{a,2-2}$ : Let  $\{x_i\}_{i=1}^p$  be i.i.d. Cauchy(0, 1). We set  $w_i = x_i + \frac{\tau_2}{p} \sum_{j \neq i} x_j$  for all  $i$ .
- $H_{a,3-1}$ : Let  $\{x_i\}_{i=1}^p$  be i.i.d.  $t(4)$ . We set  $w_1 = x_1 + \delta x_2$ ,  $w_2 = \delta x_1 + x_2$  and  $w_i = x_i$  for all  $i \neq 1, 2$ .
- $H_{a,3-2}$ : Let  $\{x_i\}_{i=1}^p$  be i.i.d.  $t(4)$ . We set  $w_i = x_i + \frac{\tau_3}{p} \sum_{j \neq i} x_j$  for all  $i$ .

Here we give more explanation on the above two types of alternatives. Let us take the Gaussian case as an example. Notice that  $A = \delta(\mathbf{e}_1 \mathbf{e}_2^* + \mathbf{e}_2 \mathbf{e}_1^*)$  is rank-two and  $B = \frac{\tau_1}{p} \mathbf{1} \mathbf{1}'$  is rank-one, where  $\mathbf{1}$  represents the all-one vector. It is easy to see that the two nonzero eigenvalues of  $A$  are  $\delta$  and  $-\delta$ , while the nonzero eigenvalue of  $B$  is  $\tau_1$ . Hence, the population covariance matrix  $I_p + A$  (resp.  $I_p + B$ ) has a spike with strength  $1 + \delta$  (resp.,  $1 + \tau_1$ ). Since the seminal work of Baik, Ben-Arous and P ech e [2], it is now well known that there is a phase transition called BBP-transition for the largest eigenvalue of the sample covariance matrix when the population covariance matrix has a spike. Very roughly speaking, we can effectively detect the spike using the largest eigenvalue of the sample covariance matrix, only when the spike is larger than the threshold  $1 + \sqrt{\frac{p}{n}}$ . Although here we are considering correlation type of matrices, simulation shows that there is a similar effect. Further, although there is no concept of population covariance matrix for Cauchy(0, 1) and  $t(4)$  variables, the alternatives  $H_{a,i-1}$  and  $H_{a,i-2}$  for  $i = 2, 3$  are constructed in a similar vein.

The results of sizes and powers stated in Table 1 are obtained under the choices  $p = 200, 400, 560, 800$  with the same  $n = 600$ . The results are based on 1000 replications. The parameters are chosen to be  $\delta = 1$ ,  $\tau_1 = \tau_3 = \frac{3}{2}$  and  $\tau_2 = \frac{1}{40}$ . We also refer to Tables S.1 and S.2 in the Supplementary Material [7] for the results under different choices of  $p$  and  $n$ . In addition, we depict the powers for different choices of the parameters  $\delta, \tau_1, \tau_2, \tau_3$  in Figures S.1–S.6 in [7], under the setting  $(p, n) = (400, 600)$ .

TABLE 1

The sizes and powers (percentage) of  $T_1$  to  $T_7$  under different hypotheses and dimension  $p$ . Here we chose sample size  $n = 600$ ,  $\delta = 1$ ,  $\tau_1 = \tau_3 = \frac{3}{2}$  and  $\tau_2 = \frac{1}{40}$

$p$	$T_1$	$T_2$	$T_3$	$T_4$	$T_5$	$T_6$	$T_7$	$T_2$	$T_4$	$T_6$	$T_7$	$T_2$	$T_4$	$T_6$	$T_7$
	$H_{0,1}$							$H_{0,2}$				$H_{0,3}$			
200	3.7	5.4	2.6	3.8	0.5	1.5	1.6	4.6	3.9	1.5	1.7	6	3.2	2.3	2.3
400	2.2	4.9	2.9	2.5	1.8	3.1	3.6	5.1	4.3	2.3	2.9	4.9	3.2	1.7	2.5
560	1.6	5.2	2.9	3.5	1.8	2.2	2.8	4.2	5	1.6	2.5	5.5	5.2	1.5	1.8
800	1.3	5.4	4.2	5	1.7	2.1	2.4	5.5	4.2	2.1	2.6	4	3.4	3.1	3.6
	$H_{a,1-1}$							$H_{a,2-1}$				$H_{a,3-1}$			
200	87.3	90.9	100	100	99.4	99.7	100	92	100	99.9	100	90.6	100	99.6	100
400	31	44	100	100	38.1	40.9	99.8	44.3	100	38.8	99.8	42.2	100	36.2	99.8
	$H_{a,1-2}$							$H_{a,2-2}$				$H_{a,3-2}$			
200	89.7	99.5	5	5.7	100	100	100	93.3	8.5	95.5	95.7	100	6.1	100	100
400	39.7	69.2	3.5	3.6	99.3	97.1	97.6	97.2	9.9	98.7	98.8	91.9	4.2	100	100
560	18.5	44.2	3.3	4.3	89.7	83.3	84.9	98.5	8.2	99.5	99.5	70	5.3	99.4	99.5
800	8.1	24.7	3	4.9	55.6	47	49.1	99.6	8.4	100	100	47	3.7	88.7	89.7

Since  $T_1$ ,  $T_3$  and  $T_5$  are parametric and the limiting theorems of them in [5, 17, 19, 30] do not apply to the Cauchy(0, 1) and  $t(4)$  variables, we omit the simulation results from the tables in these cases. Observe that for the first type of alternatives  $H_{a,i-1}$  for  $i = 1, 2, 3$ , we only consider the case when  $p$  is sufficiently smaller than  $n$ . We take  $H_{a,1-1}$  to explain such a choice. In  $H_{a,1-1}$ , we consider a Gaussian vector with a population covariance matrix  $I_p + A$ . On one hand,  $\delta$  has to be no larger than 1 to guarantee the nonnegative definiteness of  $I_p + A$ . On the other hand, as we mentioned previously, heuristically, due to the BBP transition, one needs  $\delta > \sqrt{\frac{p}{n}}$  to get effective information about the existence of  $\delta$  from the largest eigenvalue of the sample covariance matrix. Hence, in case that  $p$  is close to or larger than  $n$ , our spike  $1 + \delta$  would not be large enough to be detected. Simulation shows that a similar effect exists for all three types of correlation matrices considered here. So we omit the simulation results in those regimes where all the largest eigenvalue statistics will essentially fail.

Below we summarize our findings from the simulation study.

(1) From Table 1, and also Table S.1 and Table S.2 in the Supplementary Material [7], we see that the sizes of  $T_2$  are close to the nominal size 5%. The sizes of all the other statistics tend to be smaller than 5%. However, for the statistics of the largest eigenvalue  $T_5$ ,  $T_6$  and  $T_7$ , it is possible to modify the centering and scaling constants for the largest eigenvalues to improve the convergence rate of the weak convergence to the  $TW_1$  law such that better sizes can be achieved. Some important works have been done along this line, but only for Gaussian ensembles; see [11, 22, 27]. The extension of the results in [11, 22, 27] to other random matrix ensembles is still an open question. We do not pursue this direction in the current paper.

(2) From Table 1, and also Table S.1 and Table S.2 in the Supplementary Material [7], we see that the statistics of the largest off-diagonal entry, that is,  $T_3$ ,  $T_4$ , outperform the other statistics in the case of one large disturbance ( $H_{a,i-1}$ ,  $i = 1, 2, 3$ ). However,  $T_3$ ,  $T_4$  perform quite poorly in the case of many small disturbances ( $H_{a,i-2}$ ,  $i = 1, 2, 3$ ). In general, the other statistics perform well in both types of alternatives. In addition,  $T_7$  outperforms the others in most of the cases. For all statistics, the performance deteriorates when  $\frac{p}{n}$  increases. That can be again understood as an effect of the BBP transition. We also refer to Figures S.1–S.6 in [7] for more information about the powers for different choices of the parameters.

(3) In the Supplementary Material [7], we also consider another type of alternative hypothesis, denoted by  $H_{a,4}$ . For this alternative hypothesis, we consider a random vector  $\mathbf{w}$  which has uncorrelated but dependent components. We refer to [7] for the detailed definition. The simulation results are stated in Table S.3. One can see that  $T_4$  and  $T_7$  outperform the other statistics in general.

Overall, our statistic  $T_7$  has the following advantages. First, it is nonparametric and thus can be used for the heavy-tailed variables, for which  $T_1$ ,  $T_3$  and  $T_5$  cannot be applied. Second, among all nonparametric statistics  $T_2$ ,  $T_4$ ,  $T_6$  and  $T_7$ , only  $T_2$

performs better than  $T_7$  for the first type of alternatives, but  $T_2$  completely fails for the second type of alternatives. In a nutshell,  $T_7$  is the most robust among all 7 statistics for the cases considered in this simulation study.

**3. Hoeffding decomposition and large deviation.** In this section we state some key large deviation estimates; see Propositions 3.1 and 3.2. We start with (a variant of) Hoeffding decomposition for  $v_{k,(ij)}$ 's.

3.1. *Hoeffding decomposition.* Let

$$(3.1) \quad v_{k,(i\cdot)} := \mathbb{E}(\text{sign}(w_{ki} - w_{kj})|w_{ki}), \quad v_{k,(j\cdot)} := \mathbb{E}(\text{sign}(w_{ki} - w_{kj})|w_{kj}).$$

Observe that  $v_{k,(i\cdot)} = -v_{k,(i\cdot)}$ . The following decomposition is (a variant of) Hoeffding decomposition

$$(3.2) \quad v_{k,(ij)} = v_{k,(i\cdot)} - v_{k,(j\cdot)} + \bar{v}_{k,(ij)},$$

where we take (3.2) as the definition of  $\bar{v}_{k,(ij)}$ . It is easy to check that the three parts in the RHS are pairwise uncorrelated. In addition, all of the three parts in the RHS of (3.2) are with mean 0 and variance  $\frac{1}{3}$ , that is,

$$(3.3) \quad \mathbb{E}v_{k,(i\cdot)} = \mathbb{E}v_{k,(j\cdot)} = \mathbb{E}\bar{v}_{k,(ij)} = 0, \quad \mathbb{E}v_{k,(i\cdot)}^2 = \mathbb{E}v_{k,(j\cdot)}^2 = \mathbb{E}\bar{v}_{k,(ij)}^2 = \frac{1}{3}.$$

For brevity, we further introduce the notation

$$(3.4) \quad u_{k,(ij)} := v_{k,(i\cdot)} - v_{k,(j\cdot)}.$$

Hence, we can also write  $v_{k,(ij)} = u_{k,(ij)} + \bar{v}_{k,(ij)}$ .

For a fixed  $k \in \llbracket 1, p \rrbracket$ , let  $F_k$  be the common distribution of all  $w_{ki}$ ,  $i \in \llbracket 1, n \rrbracket$ . We see that

$$(3.5) \quad v_{k,(i\cdot)} = \mathbb{E}(\mathbb{1}(w_{kj} \leq w_{ki})|w_{ki}) - \mathbb{E}(\mathbb{1}(w_{kj} > w_{ki})|w_{ki}) = 2F_k(w_{ki}) - 1,$$

which is uniformly distributed on  $[-1, 1]$ . Hence, all  $v_{k,(i\cdot)}$ ,  $(k, i) \in \llbracket 1, p \rrbracket \times \llbracket 1, n \rrbracket$  are i.i.d., uniform random variables on  $[-1, 1]$ , in light of (3.5) and the independence of  $w_{ki}$ 's. We will call  $v_{k,(i\cdot)}$  and  $v_{k,(j\cdot)}$  (or together  $u_{k,(ij)}$ ) the *linear* parts of  $v_{k,(ij)}$ , and call  $\bar{v}_{k,(ij)}$  the *nonlinear* part. Although the linear parts in all  $v_{k,(ij)}$ 's have a simple dependence structure due to the independence between  $v_{k,(i\cdot)}$ 's, the nonlinear parts couple  $v_{k,(ij)}$ 's together with certain nontrivial dependence relation. For instance,  $v_{k,(ij)}$  and  $v_{k,(i\ell)}$  are correlated even when  $j \neq \ell$ . More specifically, it is elementary to check

$$(3.6) \quad \mathbb{E}v_{k,(ij)}v_{k,(i\ell)} = \mathbb{E}(v_{k,(i\cdot)})^2 = \frac{1}{3}.$$

In the sequel, we will often separate the nonlinear part from the linear part. To this end, we introduce the following notation. We set the  $M$ -dimensional row vector

$$\begin{aligned}
 \mathbf{v}_k &:= \frac{1}{\sqrt{M}}(v_{k,(ij)})_{i < j} \\
 (3.7) \quad &\equiv \frac{1}{\sqrt{M}}(v_{k,(12)}, \dots, v_{k,(1n)}, v_{k,(23)}, \dots, v_{k,(2n)}, \dots, v_{k,(n-1,n)}).
 \end{aligned}$$

Further, we set

$$(3.8) \quad \mathbf{u}_k := \frac{1}{\sqrt{M}}(u_{k,(ij)})_{i < j}, \quad \bar{\mathbf{v}}_k := \frac{1}{\sqrt{M}}(\bar{v}_{k,(ij)})_{i < j}.$$

With the above notation, we can write

$$(3.9) \quad \mathbf{v}_k = \mathbf{u}_k + \bar{\mathbf{v}}_k, \quad k \in \llbracket 1, p \rrbracket.$$

Note that under the null hypothesis, that is, the components of the population vector  $\mathbf{w}$  are independent, the random vectors  $\mathbf{v}_1, \dots, \mathbf{v}_p$  are also independent. But the components in  $\mathbf{v}_k$  are dependent, as mentioned above (cf. (3.6)). We also notice that  $\mathbf{v}_i$  is the  $i$ th row of  $\Theta$  defined in (1.5). For the columns of  $\Theta$ , that is,  $\boldsymbol{\theta}_{(ij)}$ 's in (1.3), we also introduce the notation

$$\boldsymbol{\theta}_{(i\cdot)} := \frac{1}{\sqrt{M}}(v_{1,(i\cdot)}, \dots, v_{p,(i\cdot)})', \quad \bar{\boldsymbol{\theta}}_{(ij)} := \frac{1}{\sqrt{M}}(\bar{v}_{1,(ij)}, \dots, \bar{v}_{p,(ij)})'.$$

Hence, we have the decomposition for columns

$$(3.10) \quad \boldsymbol{\theta}_{(ij)} = \boldsymbol{\theta}_{(i\cdot)} - \boldsymbol{\theta}_{(j\cdot)} + \bar{\boldsymbol{\theta}}_{(ij)}.$$

Further note that the nonzero eigenvalues of the matrix  $K$  are the same as those of the following  $M \times M$  matrix:

$$(3.11) \quad \mathcal{K} := \sum_{i=1}^p \mathbf{v}'_i \mathbf{v}_i = \Theta' \Theta.$$

3.2. *Large deviation estimates for  $\mathbf{v}_k$ .* Set the  $M \times M$  symmetric matrix

$$(3.12) \quad \Gamma = (\chi_{(ij)(st)})_{i < j, s < t},$$

where  $(ij)$  is the row index and  $(st)$  is the column index and

$$\chi_{(ij)(st)} := \frac{1}{3}(\delta_{is} + \delta_{jt} - \delta_{it} - \delta_{js}).$$

It is elementary to check that

$$(3.13) \quad \Gamma^2 = \frac{n}{3}\Gamma.$$

Consequently, we have the fact

$$(3.14) \quad \|\Gamma\| = O(n).$$

We further set the  $n \times M$  matrix

$$(3.15) \quad T = (t_{\ell,(ij)})_{\ell,i < j}, \quad t_{\ell,(ij)} := \delta_{\ell i} - \delta_{\ell j}, \quad 1 \leq \ell \leq n, 1 \leq i < j \leq n,$$

where  $\ell$  is the row index and  $(ij)$  is the column index. It is easy to check

$$(3.16) \quad \Gamma = \frac{1}{3}T'T.$$

The first proposition is on the large deviation estimates for some linear and quadratic forms of  $\mathbf{u}_k$ .

**PROPOSITION 3.1.** *Let  $\mathbf{u}_k$  be defined as in (3.8). Let  $\mathbf{a} = (a_{(ij)})_{i < j} \in \mathbb{C}^M$  be any deterministic vector, and let  $B := (b_{(ij),(st)})_{i < j, s < t} \in \mathbb{C}^{M \times M}$  be any deterministic matrix. We have*

$$(3.17) \quad \mathbb{E} \mathbf{u}_k B \mathbf{u}'_k = \frac{1}{M} \text{Tr} B \Gamma,$$

$$(3.18) \quad |\mathbf{u}_k \mathbf{a}'| < \sqrt{\frac{\mathbf{a} \Gamma \mathbf{a}^*}{M}} < \sqrt{\frac{\|\mathbf{a}\|^2}{n}},$$

$$(3.19) \quad \left| \mathbf{u}_k B \mathbf{u}'_k - \frac{1}{M} \text{Tr} B \Gamma \right| < \sqrt{\frac{\text{Tr} |B \Gamma|^2}{M^2}}.$$

The second proposition is about the large deviation estimates for some linear and quadratic forms of  $\bar{\mathbf{v}}_k$  and the crossing quadratic forms of  $\bar{\mathbf{v}}_k$  and  $\mathbf{u}_k$ .

**PROPOSITION 3.2.** *Let  $\mathbf{u}_k$  and  $\bar{\mathbf{v}}_k$  be as defined in (3.8). Let  $\mathbf{a} = (a_{(ij)})_{i < j} \in \mathbb{C}^M$  be any deterministic vector, and let  $B := (b_{(ij),(st)})_{i < j, s < t} \in \mathbb{C}^{M \times M}$  be any deterministic matrix. We have*

$$(3.20) \quad |\bar{\mathbf{v}}_k \mathbf{a}'| < \sqrt{\frac{\|\mathbf{a}\|^2}{M}},$$

$$(3.21) \quad |\mathbf{u}_k B \bar{\mathbf{v}}'_k| < \sqrt{\frac{n}{M^2} \text{Tr} |B|^2} + \sqrt{\frac{1}{M^2} \sum_{\ell=1}^n \left| \sum_{j=\ell+1}^n (TB)_{j,(\ell j)} \right|^2},$$

$$(3.22) \quad \left| \bar{\mathbf{v}}_k B \bar{\mathbf{v}}'_k - \frac{1}{3M} \text{Tr} B \right| < \sqrt{\frac{n}{M^2} \text{Tr} |B|^2}.$$

We further set

$$(3.23) \quad \tilde{\Gamma} = \Gamma + \frac{1}{3}I_M.$$

From Propositions 3.1 and 3.2, we can easily get the following corollary.

**COROLLARY 3.3.** *Let  $\mathbf{v}_k$  be as defined in (3.7). Let  $\mathbf{a} = (a_{(ij)})_{i < j} \in \mathbb{C}^M$  be any deterministic vector, and let  $B := (b_{(ij),(st)})_{i < j, s < t} \in \mathbb{C}^{M \times M}$  be any deterministic matrix. We have*

$$(3.24) \quad |\mathbf{v}_k \mathbf{a}'| \prec \sqrt{\frac{\mathbf{a} \Gamma \mathbf{a}'}{M}} \prec \sqrt{\frac{\|\mathbf{a}\|^2}{n}},$$

$$(3.25) \quad \left| \mathbf{v}_k B \mathbf{v}'_k - \frac{1}{M} \text{Tr} B \tilde{\Gamma} \right| \prec \sqrt{\frac{\text{Tr} |B|^2}{M}}.$$

The proofs of Propositions 3.1 and 3.2 and also the proof of Corollary 3.3 are stated in the Supplementary Material [7].

**4. Strong local law for  $K$ .** In this section we state a strong local law for the matrix  $K$ ; see Proposition 4.1. The proof of Proposition 4.1 is stated in the Supplementary Material [7] and it heavily relies on the large deviation bounds in Corollary 3.3. To state the results, we need more notation. Recall the matrices  $K$  and  $\mathcal{K}$  defined in (1.4) and (3.11). We denote the Green functions of  $K$  and  $\mathcal{K}$  by

$$G(z) := (K - z)^{-1}, \quad \mathcal{G}(z) := (\mathcal{K} - z)^{-1}.$$

Then, we further denote the Stieltjes transform of  $K$  by

$$m(z) := \frac{1}{p} \text{Tr} G(z) = \frac{1}{p} \sum_{i=1}^p G_{ii}(z).$$

For any  $z = E + i\eta \in \mathbb{C}^+$ , we set the function  $\underline{m}(z) : \mathbb{C}^+ \rightarrow \mathbb{C}^+$  as the solution to the equation

$$(4.1) \quad \frac{2}{3}c_n \left( z - \frac{1}{3} \right) (\underline{m}(z))^2 + \left( z - 1 + \frac{2}{3}c_n \right) \underline{m}(z) + 1 = 0.$$

It is elementary to check that  $\underline{m}$  is the Stieltjes transform of  $F_{c_n}^K$  (cf. Theorem 1.1). Some properties of the function  $\underline{m}$  are given in Lemma S0.5.

We then introduce the following notation:

$$(4.2) \quad \begin{aligned} \Lambda_d \equiv \Lambda_d(z) &:= \max_k |G_{kk}(z) - \underline{m}(z)|, & \Lambda_o \equiv \Lambda_o(z) &:= \max_{k \neq \ell} |G_{k\ell}(z)|, \\ \Lambda \equiv \Lambda(z) &:= |m(z) - \underline{m}(z)|. \end{aligned}$$

In the sequel, we work in the following domain of  $z$ :

$$(4.3) \quad \mathcal{D}(\epsilon) := \left\{ z = E + i\eta : \frac{1}{2}\lambda_{+,c} \leq E \leq 2\lambda_{+,c}, n^{-1+\epsilon} \leq \eta \leq 1 \right\},$$

where  $\lambda_{+,c}$  is defined in (1.6). Let  $\gamma_1 \geq \gamma_2 \geq \dots \geq \gamma_{p \wedge n}$  be the ordered  $p$ -quantiles of  $F_{c_n}^K$ , that is,  $\gamma_j$  is the smallest real number such that

$$\int_{-\infty}^{\gamma_j} dF_{c_n}^K(x) = \frac{p - j + 1}{p}, \quad j \in [1, n \wedge p].$$

We further define the deterministic control parameter

$$(4.4) \quad \Psi \equiv \Psi(z) := \sqrt{\frac{\text{Im } m(z)}{n\eta}} + \frac{1}{n\eta}.$$

With the above notation we can now state the following strong local law.

PROPOSITION 4.1. *Under assumption (1.1) the following hold:*

(i) (Entrywise local law). *The following bounds hold uniformly on  $\mathcal{D}(\epsilon)$ :*

$$(4.5) \quad \Lambda_d(z) \prec \Psi(z), \quad \Lambda_o(z) \prec \Psi(z).$$

(ii) (Strong local law). *The following bound holds uniformly on  $\mathcal{D}(\epsilon)$ :*

$$(4.6) \quad \Lambda(z) \prec \frac{1}{n\eta}.$$

(iii) (Rigidity on the right edge). *For  $i \in [1, \delta p]$  with any sufficiently small constant  $\delta \in (0, 1)$ , we have*

$$(4.7) \quad |\lambda_i(K) - \gamma_i| \prec n^{-\frac{2}{3}i - \frac{1}{3}}.$$

**5. Decoupling.** In this section we compare the Green functions of the matrix  $K$  with another random matrix  $\widehat{K}$  which has independent linear part and “nonlinear” part (cf. (5.2)). Recall (3.1). We set the matrices

$$(5.1) \quad U := \frac{1}{\sqrt{M}}((v_{k,(i\cdot)} - v_{k,(j\cdot)}))_{k,(ij)}, \quad \bar{V} := \frac{1}{\sqrt{M}}(\bar{v}_{k,(ij)})_{k,(ij)}$$

and let

$$H := \frac{1}{\sqrt{M}}(h_{k,(ij)})_{k,(ij)}, \quad k \in \llbracket 1, p \rrbracket, 1 \leq i < j \leq n$$

be a  $p \times M$  matrix, where the entries  $h_{k,(ij)}$ ’s are i.i.d.  $N(0, \frac{1}{3})$ . We also set the random variables  $h_{k,(ij)} := -h_{k,(ji)}$  if  $i \geq j$ , for further use. We assume that  $H$  is independent of  $U$ . We define the random matrices

$$(5.2) \quad \widehat{\Theta} := (U + H), \quad \widehat{K} := \widehat{\Theta}\widehat{\Theta}' = (U + H)(U + H)'$$

Then we denote the Green function of  $\widehat{K}$  and its normalized trace by

$$\widehat{G}(z) := (\widehat{K} - z)^{-1}, \quad \widehat{m}(z) := \frac{1}{p} \text{Tr } \widehat{G}(z).$$

In this section we will establish the following comparison proposition.

PROPOSITION 5.1. *Let  $\varepsilon > 0$  be any sufficiently small constant. Set  $\eta = n^{-\frac{2}{3}-\varepsilon}$ . Let  $E_1, E_2 \in \mathbb{R}$  satisfy  $E_1 < E_2$  and*

$$(5.3) \quad |E_1|, |E_2| \leq n^{-\frac{2}{3}+\varepsilon}.$$

*Let  $F : \mathbb{R} \rightarrow \mathbb{R}$  be a smooth function satisfying  $\max_{x \in \mathbb{R}} |F^{(\ell)}(x)|(|x| + 1)^{-C} \leq C$ ,  $\ell = 1, 2, 3, 4$ , for some positive constant  $C$ . Then there exists a constant  $\delta > 0$  such that, for sufficiently large  $n$  we have*

$$\left| \mathbb{E}F\left(n \int_{E_1}^{E_2} \operatorname{Im} m(x + \lambda_{+,c_n} + i\eta) \, dx\right) - \mathbb{E}F\left(n \int_{E_1}^{E_2} \operatorname{Im} \widehat{m}(x + \lambda_{+,c_n} + i\eta) \, dx\right) \right| \leq n^{-\delta}.$$

PROOF. For simplicity, in this proof, we denote by

$$(5.4) \quad z \equiv z(x) := x + \lambda_{+,c_n} + i\eta, \quad x \in [E_1, E_2].$$

Recall the small constant  $\varepsilon$  in Proposition 5.1. For brevity, we will simply write  $C\varepsilon$  with any positive constant (independent of  $\varepsilon$ ) by  $\varepsilon$  in the sequel. In other words, we allow  $\varepsilon$  to vary from line to line, up to  $C$ . We then construct the following sequence of the interpolations:  $\Theta = \Theta_0, \dots, \Theta_{\gamma-1}, \Theta_\gamma, \dots, \Theta_p = \widehat{\Theta}$ , where  $\Theta_\gamma$  is the matrix whose first  $\gamma$  rows are the same as those of  $\widehat{\Theta}$  and the remaining  $p - \gamma$  rows are the same as those of  $\Theta$ . Correspondingly, we set the notation

$$K_\gamma = \Theta_\gamma \Theta'_\gamma, \quad G_\gamma(z) := (K_\gamma - z)^{-1}, \quad m_\gamma := \frac{1}{p} \operatorname{Tr} G_\gamma(z).$$

We first claim the following lemma, whose proof is stated in the Supplementary Material [7].

LEMMA 5.2 (Local law for  $K_\gamma$ ). *All the estimates in Proposition 4.1 hold for  $K_\gamma$  for all  $\gamma \in \llbracket 0, p \rrbracket$ .*

With Lemma 5.2, we proceed to the proof of Proposition 5.1. Using the above notation, we can write

$$\begin{aligned} & \mathbb{E}F\left(n \int_{E_1}^{E_2} \operatorname{Im} m(z) \, dx\right) - \mathbb{E}F\left(n \int_{E_1}^{E_2} \operatorname{Im} \widehat{m}(z) \, dx\right) \\ &= \mathbb{E}F\left(n \int_{E_1}^{E_2} \operatorname{Im} m_0(z) \, dx\right) - \mathbb{E}F\left(n \int_{E_1}^{E_2} \operatorname{Im} m_p(z) \, dx\right) \\ &= \sum_{\gamma=1}^p \left( \mathbb{E}F\left(n \int_{E_1}^{E_2} \operatorname{Im} m_{\gamma-1}(z) \, dx\right) - \mathbb{E}F\left(n \int_{E_1}^{E_2} \operatorname{Im} m_\gamma(z) \, dx\right) \right). \end{aligned}$$

Hence, it suffices to show that for all  $\gamma \in \llbracket 1, p \rrbracket$ ,

$$(5.5) \quad \left| \mathbb{E}F \left( n \int_{E_1}^{E_2} \text{Im} m_{\gamma-1}(z) \, dx \right) - \mathbb{E}F \left( n \int_{E_1}^{E_2} \text{Im} m_\gamma(z) \, dx \right) \right| \leq n^{-1-\delta}$$

for some positive constant  $\delta$ . For a fixed  $\gamma$ , we further introduce the notation  $\Theta_\gamma^{(i)}$  to denote the matrix obtained from  $\Theta_\gamma$  with the  $i$ th row removed. Then, by definition, we have  $\Theta_{\gamma-1}^{(\gamma)} = \Theta_\gamma^{(\gamma)}$ . Correspondingly, we use the notation

$$K_\gamma^{(i)} := \Theta_\gamma^{(i)} (\Theta_\gamma^{(i)})', \quad G_\gamma^{(i)} := (K_\gamma^{(i)} - z)^{-1}, \quad m_\gamma^{(i)} := \frac{1}{p} \text{Tr} G_\gamma^{(i)}.$$

Also note that  $m_{\gamma-1}^{(\gamma)} = m_\gamma^{(\gamma)}$ . Next we expand both  $m_{\gamma-1}$  and  $m_\gamma$  around  $m_\gamma^{(\gamma)}$ . Observe that

$$(5.6) \quad \begin{aligned} m_{\gamma-1} - m_\gamma^{(\gamma)} &= \frac{1}{p} \frac{1 + \mathbf{v}_\gamma (\Theta_\gamma^{(\gamma)})' (G_\gamma^{(\gamma)})^2 \Theta_\gamma^{(\gamma)} \mathbf{v}'_\gamma}{\mathbf{v}_\gamma \mathbf{v}'_\gamma - z - \mathbf{v}_\gamma (\Theta_\gamma^{(\gamma)})' G_\gamma^{(\gamma)} \Theta_\gamma^{(\gamma)} \mathbf{v}'_\gamma} \\ &=: \frac{1}{p} \frac{1 + \mathbf{v}_\gamma A_\gamma \mathbf{v}'_\gamma}{1 - z - \mathbf{v}_\gamma B_\gamma \mathbf{v}'_\gamma}, \end{aligned}$$

where in the last step we use the trivial fact  $\mathbf{v}_\gamma \mathbf{v}'_\gamma = 1$ . Similarly,

$$(5.7) \quad m_\gamma - m_\gamma^{(\gamma)} = \frac{1}{p} \frac{1 + \hat{\mathbf{v}}_\gamma A_\gamma \hat{\mathbf{v}}'_\gamma}{\hat{\mathbf{v}}_\gamma \hat{\mathbf{v}}'_\gamma - z - \hat{\mathbf{v}}_\gamma B_\gamma \hat{\mathbf{v}}'_\gamma},$$

where we use the notation  $\hat{\mathbf{v}}_\gamma := \mathbf{u}_\gamma + \mathbf{h}_\gamma$  to denote the  $\gamma$ th row of  $\hat{\Theta}$ .

We then further set

$$(5.8) \quad \begin{aligned} D_\gamma &:= \mathbf{v}_\gamma B_\gamma \mathbf{v}'_\gamma - \frac{1}{M} \text{Tr} B_\gamma \Gamma, \\ \hat{D}_\gamma &:= 1 - \hat{\mathbf{v}}_\gamma \hat{\mathbf{v}}'_\gamma + \hat{\mathbf{v}}_\gamma B_\gamma \hat{\mathbf{v}}'_\gamma - \frac{1}{M} \text{Tr} B_\gamma \Gamma, \end{aligned}$$

and write

$$(5.9) \quad \begin{aligned} D_\gamma &= \left( \mathbf{u}_\gamma B_\gamma \mathbf{u}'_\gamma - \frac{1}{M} \text{Tr} B_\gamma \Gamma \right) + \bar{\mathbf{v}}_\gamma B_\gamma \bar{\mathbf{v}}'_\gamma + 2\mathbf{u}_\gamma B_\gamma \bar{\mathbf{v}}'_\gamma \\ &=: \mathcal{U}_\gamma + \mathcal{V}_\gamma + \mathcal{P}_\gamma, \\ \hat{D}_\gamma &= \left( \mathbf{u}_\gamma B_\gamma \mathbf{u}'_\gamma - \frac{1}{M} \text{Tr} B_\gamma \Gamma \right) + \mathbf{h}_\gamma B_\gamma \mathbf{h}'_\gamma + 2\mathbf{u}_\gamma B_\gamma \mathbf{h}'_\gamma \\ &\quad + \left( \frac{2}{3} - \mathbf{u}_\gamma \mathbf{u}'_\gamma \right) + \left( \frac{1}{3} - \mathbf{h}_\gamma \mathbf{h}'_\gamma - 2\mathbf{u}_\gamma \mathbf{h}'_\gamma \right) \\ &=: \mathcal{U}_\gamma + \hat{\mathcal{V}}_\gamma + \hat{\mathcal{P}}_\gamma + \hat{\mathcal{W}}_\gamma + \hat{\mathcal{O}}_\gamma, \end{aligned}$$

where we recall that  $B_\gamma$  is (complex) symmetric. Similarly, we write

$$\begin{aligned}
 \mathbf{v}_\gamma A_\gamma \mathbf{v}'_\gamma &= \mathbf{u}_\gamma A_\gamma \mathbf{u}'_\gamma + \bar{\mathbf{v}}_\gamma A_\gamma \bar{\mathbf{v}}'_\gamma + 2\mathbf{u}_\gamma A_\gamma \bar{\mathbf{v}}'_\gamma \\
 &=: \mathbf{u}_\gamma A_\gamma \mathbf{u}'_\gamma + \bar{\mathbf{v}}_\gamma A_\gamma \bar{\mathbf{v}}'_\gamma + \mathcal{Q}_\gamma, \\
 \hat{\mathbf{v}}_\gamma A_\gamma \hat{\mathbf{v}}'_\gamma &= \mathbf{u}_\gamma A_\gamma \mathbf{u}'_\gamma + \mathbf{h}_\gamma A_\gamma \mathbf{h}'_\gamma + 2\mathbf{u}_\gamma A_\gamma \mathbf{h}'_\gamma \\
 &=: \mathbf{u}_\gamma A_\gamma \mathbf{u}'_\gamma + \mathbf{h}_\gamma A_\gamma \mathbf{h}'_\gamma + \hat{\mathcal{Q}}_\gamma.
 \end{aligned}
 \tag{5.10}$$

We have the following crucial technical lemma.

LEMMA 5.3. *Let  $\eta = n^{-\frac{2}{3}-\varepsilon}$ , and  $x, x_1, x_2 \in [E_1, E_2]$ , where  $E_1$  and  $E_2$  satisfy (5.3). Let  $z = x + \lambda_{+,c_n} + i\eta$  and  $z_a = x_a + \lambda_{+,c_n} + i\eta$ ,  $a = 1, 2$ . With the above notation, we have*

$$\begin{aligned}
 |\mathcal{U}_\gamma(z)| &< n^{-\frac{1}{3}+\varepsilon}, & |\mathcal{V}_\gamma(z)| &< n^{-\frac{5}{6}+\varepsilon}, & |\hat{\mathcal{P}}_\gamma(z)| &< n^{-\frac{5}{6}+\varepsilon}, \\
 |\hat{\mathcal{V}}_\gamma(z)| &< n^{-1+\varepsilon}, & |\hat{\mathcal{W}}_\gamma(z)| &< n^{-\frac{1}{2}+\varepsilon}, & |\hat{\mathcal{O}}_\gamma(z)| &< n^{-1+\varepsilon}, \\
 |\hat{\mathcal{Q}}_\gamma(z)| &< n^{-\frac{1}{6}+\varepsilon}, & |\mathbf{u}_\gamma A_\gamma(z) \mathbf{u}'_\gamma| &< n^{\frac{1}{3}+\varepsilon}, \\
 |\bar{\mathbf{v}}_\gamma A_\gamma(z) \bar{\mathbf{v}}'_\gamma| &< n^{-\frac{1}{6}+\varepsilon}, & |\mathbf{h}_\gamma A_\gamma(z) \mathbf{h}'_\gamma| &< n^{-\frac{1}{2}+\varepsilon},
 \end{aligned}
 \tag{5.11}$$

and

$$|\mathcal{P}_\gamma(z)| < n^{-\frac{1}{2}+\varepsilon}, \quad |\mathcal{Q}_\gamma(z)| < n^{-\frac{1}{6}+\varepsilon}.
 \tag{5.12}$$

In addition, we have

$$|\mathbb{E}(\mathbf{u}_\gamma A_\gamma(z) \mathbf{u}'_\gamma \hat{\mathcal{W}}_\gamma)| < n^{-\frac{2}{3}+\varepsilon}, \quad |\mathbb{E}(\mathbf{u}_\gamma A_\gamma(z_1) \mathbf{u}'_\gamma \mathcal{P}_\gamma(z_2))| < n^{-\frac{1}{2}+\varepsilon}.
 \tag{5.13}$$

The above estimates still hold if we replace some or all of  $z, z_1, z_2$  by their complex conjugates.

The proof of Lemma 5.3 will be stated in the Supplementary Material [7]. Two key technical inputs for the proof are Propositions 3.1 and 3.2.

We proceed to the proof of Proposition 5.1, with the aid of Lemma 5.3. First, using (5.11) and (5.12), we can write

$$\begin{aligned}
 n \int_{E_1}^{E_2} (m_{\gamma-1}(z) - m_\gamma^{(\gamma)}(z)) dx &= \frac{n}{p} \int_{E_1}^{E_2} \frac{1 + \mathbf{v}_\gamma A_\gamma \mathbf{v}'_\gamma}{1 - z - \frac{1}{M} \text{Tr } B_\gamma \Gamma - D_\gamma} dx \\
 &= \tau_{\gamma 0} + \tau_{\gamma 1} + \tau_{\gamma 2} + O_{<}(n^{-\frac{7}{6}+\varepsilon}),
 \end{aligned}
 \tag{5.14}$$

where

$$\begin{aligned}
 \tau_{\gamma 0} &:= \frac{n}{p} \int_{E_1}^{E_2} \frac{1 + \mathbf{v}_\gamma A_\gamma \mathbf{v}'_\gamma}{\left(1 - z - \frac{1}{M} \text{Tr } B_\gamma \Gamma\right)} dx = O_{<}(n^{-\frac{1}{3}+\varepsilon}), \\
 (5.15) \quad \tau_{\gamma 1} &:= \frac{n}{p} \int_{E_1}^{E_2} \frac{1 + \mathbf{u}_\gamma A_\gamma \mathbf{u}'_\gamma}{\left(1 - z - \frac{1}{M} \text{Tr } B_\gamma \Gamma\right)^2} (\mathcal{U}_\gamma + \mathcal{P}_\gamma) dx = O_{<}(n^{-\frac{2}{3}+\varepsilon}), \\
 \tau_{\gamma 2} &:= \frac{n}{p} \int_{E_1}^{E_2} \frac{1 + \mathbf{u}_\gamma A_\gamma \mathbf{u}'_\gamma}{\left(1 - z - \frac{1}{M} \text{Tr } B_\gamma \Gamma\right)^3} \mathcal{U}_\gamma^2 dx = O_{<}(n^{-1+\varepsilon}).
 \end{aligned}$$

Here we use the fact  $1/(1 - z - \frac{1}{M} \text{Tr } B_\gamma \Gamma) \sim 1$  with high probability, which follows from  $1/(1 - z - \frac{1}{M} \text{Tr } B_\gamma \Gamma) = \underline{m} + O_{<}(\frac{1}{n\eta})$  (cf. Lemma 5.2 and an analogue of (S.75)), and also (S.122). Analogously, we have

$$\begin{aligned}
 n \int_{E_1}^{E_2} (m_\gamma(z) - m_\gamma^{(\gamma)}(z)) dx &= \frac{n}{p} \int_{E_1}^{E_2} \frac{1 + \hat{\mathbf{v}}_\gamma A_\gamma \hat{\mathbf{v}}'_\gamma}{1 - z - \frac{1}{M} \text{Tr } B_\gamma \Gamma - \hat{D}_\gamma} dx \\
 (5.16) \qquad \qquad \qquad &= \hat{\tau}_{\gamma 0} + \hat{\tau}_{\gamma 1} + \tau_{\gamma 2} + O_{<}(n^{-\frac{7}{6}+\varepsilon}),
 \end{aligned}$$

where

$$\begin{aligned}
 \hat{\tau}_{\gamma 0} &:= \frac{n}{p} \int_{E_1}^{E_2} \frac{1 + \hat{\mathbf{v}}_\gamma A_\gamma \hat{\mathbf{v}}'_\gamma}{\left(1 - z - \frac{1}{M} \text{Tr } B_\gamma \Gamma\right)} dx = O_{<}(n^{-\frac{1}{3}+\varepsilon}), \\
 (5.17) \quad \hat{\tau}_{\gamma 1} &:= \frac{n}{p} \int_{E_1}^{E_2} \frac{1 + \mathbf{u}_\gamma A_\gamma \mathbf{u}'_\gamma}{\left(1 - z - \frac{1}{M} \text{Tr } B_\gamma \Gamma\right)^2} (\mathcal{U}_\gamma + \hat{\mathcal{W}}_\gamma) dx = O_{<}(n^{-\frac{2}{3}+\varepsilon}).
 \end{aligned}$$

For brevity, we further introduce the notation  $\zeta_\gamma := n \int_{E_1}^{E_2} \text{Im } m_\gamma^{(\gamma)}(z) dx$ . Then we can write

$$\begin{aligned}
 F\left(n \int_{E_1}^{E_2} \text{Im } m_{\gamma-1}(z) dx\right) &= F(\zeta_\gamma) + F'(\zeta_\gamma)(\text{Im } \tau_{\gamma 0} + \text{Im } \tau_{\gamma 1} + \text{Im } \tau_{\gamma 2}) \\
 &\quad + \frac{F^{(2)}(\zeta_\gamma)}{2} ((\text{Im } \tau_{\gamma 0})^2 + 2 \text{Im } \tau_{\gamma 0} \text{Im } \tau_{\gamma 1}) \\
 &\quad + \frac{F^{(3)}(\zeta_\gamma)}{6} (\text{Im } \tau_{\gamma 0})^3 + O_{<}(n^{-\frac{7}{6}+\varepsilon}).
 \end{aligned}$$

Analogously, we have

$$\begin{aligned}
 F\left(n \int_{E_1}^{E_2} \text{Im } m_\gamma(z) dx\right) &= F(\zeta_\gamma) + F'(\zeta_\gamma)(\text{Im } \hat{\tau}_{\gamma 0} + \text{Im } \hat{\tau}_{\gamma 1} + \text{Im } \tau_{\gamma 2}) \\
 &\quad + \frac{F^{(2)}(\zeta_\gamma)}{2} ((\text{Im } \hat{\tau}_{\gamma 0})^2 + 2 \text{Im } \hat{\tau}_{\gamma 0} \text{Im } \hat{\tau}_{\gamma 1}) \\
 &\quad + \frac{F^{(3)}(\zeta_\gamma)}{6} (\text{Im } \hat{\tau}_{\gamma 0})^3 + O_{<}(n^{-\frac{7}{6}+\varepsilon}).
 \end{aligned}$$

Therefore, to establish (5.5), it suffices to show the following:

$$(5.18) \quad \mathbb{E} \operatorname{Im} \tau_{\gamma a} - \mathbb{E} \operatorname{Im} \hat{\tau}_{\gamma a} = O_{\prec}(n^{-1-\delta}), \quad a = 0, 1,$$

$$(5.19) \quad \mathbb{E}(\operatorname{Im} \tau_{\gamma 0})^2 - \mathbb{E}(\operatorname{Im} \hat{\tau}_{\gamma 0})^2 = O_{\prec}(n^{-1-\delta}),$$

$$(5.20) \quad \mathbb{E} \operatorname{Im} \tau_{\gamma 0} \operatorname{Im} \tau_{\gamma 1} - \mathbb{E} \operatorname{Im} \hat{\tau}_{\gamma 0} \operatorname{Im} \hat{\tau}_{\gamma 1} = O_{\prec}(n^{-1-\delta}),$$

$$(5.21) \quad \mathbb{E}(\operatorname{Im} \tau_{\gamma 0})^3 - \mathbb{E}(\operatorname{Im} \hat{\tau}_{\gamma 0})^3 = O_{\prec}(n^{-1-\delta}).$$

We prove the above estimates one by one. First, for (5.18) with  $a = 0$ , we simply have  $\mathbb{E} \operatorname{Im} \tau_{\gamma 0} - \mathbb{E} \operatorname{Im} \hat{\tau}_{\gamma 0} = 0$ , since the covariance matrix of  $\mathbf{v}_{\gamma}$  and that of  $\hat{\mathbf{v}}_{\gamma}$  are the same. For (5.18) with  $a = 1$ , the conclusion follows from the estimates in (5.13) and the bounds of  $\mathcal{P}_{\gamma}$  and  $\hat{\mathcal{W}}_{\gamma}$  in (5.11).

Next we show (5.19). Observe that for any  $\omega_1, \omega_2 \in \mathbb{C}$ , we can write  $\operatorname{Im} \omega_1 \operatorname{Im} \omega_2 = \frac{1}{4}(\omega_1 \bar{\omega}_2 + \bar{\omega}_1 \omega_2 - \omega_1 \omega_2 - \bar{\omega}_1 \bar{\omega}_2)$ . According to the definitions in (5.15) and (5.17), and also the fact that the covariance matrix of  $\mathbf{v}_{\gamma}$  and that of  $\hat{\mathbf{v}}_{\gamma}$  are the same, it suffices to show

$$(5.22) \quad \mathbb{E} \mathbf{v}_{\gamma} A_{\gamma}(z_1) \mathbf{v}'_{\gamma} \mathbf{v}_{\gamma} A_{\gamma}(z_2) \mathbf{v}'_{\gamma} - \mathbb{E} \hat{\mathbf{v}}_{\gamma} A_{\gamma}(z_1) \hat{\mathbf{v}}'_{\gamma} \hat{\mathbf{v}}_{\gamma} A_{\gamma}(z_2) \hat{\mathbf{v}}'_{\gamma} = O_{\prec}(n^{\frac{1}{3}-\delta}),$$

and, if we replace one or both of  $z_1$  and  $z_2$  by their complex conjugates, the analogues of (5.22) are also true. Here  $z_1, z_2$  satisfy the assumptions in Lemma 5.3. These desired estimates follow from the decompositions in (5.10), and the bounds in (5.11) for the terms in the decompositions. Similarly, applying the decompositions in (5.10), and the bounds in (5.11) again, one can show (5.20) and (5.21). We omit the details. This completes the proof of Proposition 5.1.  $\square$

**6. First-order approximation.** Recall (5.1). We first set

$$(6.1) \quad \tilde{K} := \frac{1}{3} I_p + U U', \quad \tilde{G}(z) := (\tilde{K} - z)^{-1}, \quad \tilde{m}(z) := \frac{1}{p} \operatorname{Tr} \tilde{G}(z).$$

In this section our aim is to establish the following proposition.

**PROPOSITION 6.1.** *Suppose that the assumptions on  $\eta, E_1, E_2, F$  in Proposition 5.1 hold. For some constant  $\delta > 0$  and sufficiently large  $n$ , we have*

$$\left| \mathbb{E} F \left( n \int_{E_1}^{E_2} \operatorname{Im} \hat{m}(x + \lambda_{+, c_n} + i\eta) dx \right) - \mathbb{E} F \left( n \int_{E_1}^{E_2} \operatorname{Im} \tilde{m}(x + \lambda_{+, c_n} + i\eta) dx \right) \right| \leq n^{-\delta}.$$

**PROOF.** We first define the following continuous interpolation between  $\hat{K}$  and  $\tilde{K}$  and its Green function for  $t \in [0, 1]$ ,

$$(6.2) \quad \hat{K}_t := (U + tH)(U + tH)' + \frac{1}{3}(1 - t^2)I_p, \quad \hat{G}_t := (\hat{K}_t - z)^{-1}.$$

and we also denote by  $\widehat{m}_t := \frac{1}{p} \text{Tr} \widehat{G}_t$ . Especially, we have  $\widehat{K}_1 = \widehat{K}$  and  $\widehat{K}_0 = \widetilde{K}$ . Similar to Lemma 5.2, we have the following local law for  $\widehat{K}_t$ , whose proof is stated in the Supplementary Material [7].

LEMMA 6.2 (Local law for  $\widehat{K}_t$ ). *All the estimates in Proposition 4.1 hold for  $\widehat{K}_t$  for all  $t \in [0, 1]$ .*

With the aid of Lemma 6.2, we now proceed to the proof of Proposition 6.1. For brevity, we simply write  $z \equiv z(x) := x + \lambda_{+,c_n} + i\eta$  in the sequel, and further introduce the notation

$$(6.3) \quad \Phi_t := n \int_{E_1}^{E_2} \text{Im} \widehat{m}_t(z) \, dx.$$

Then we can write

$$\begin{aligned} & \mathbb{E}F\left(n \int_{E_1}^{E_2} \text{Im} \widehat{m}(z) \, dx\right) - \mathbb{E}F\left(n \int_{E_1}^{E_2} \text{Im} \widetilde{m}(z) \, dx\right) \\ &= \int_0^1 \mathbb{E} \frac{\partial}{\partial t} F(\Phi_t) \, dt = \int_0^1 \mathbb{E} \left( F'(\Phi_t) \frac{\partial \Phi_t}{\partial t} \right) \, dt. \end{aligned}$$

Our aim is to show

$$\left| \frac{\partial \Phi_t}{\partial t} \right| < n^{-\delta} \quad \forall t \in [0, 1].$$

This, together with the assumption on  $F'$ , leads to the conclusions in Proposition 6.1. From the definition in (6.3) we have

$$\frac{\partial \Phi_t}{\partial t} = n \int_{E_1}^{E_2} \frac{\partial \text{Im} \widehat{m}_t(z)}{\partial t} \, dx = \frac{n}{p} \int_{E_1}^{E_2} \frac{\partial \text{Im} \text{Tr} \widehat{G}_t(z)}{\partial t} \, dx.$$

Considering that  $|E_1|, |E_2| \leq N^{-\frac{2}{3}+\varepsilon}$ , it suffices to show

$$(6.4) \quad \left| \frac{\partial \text{Tr} \widehat{G}_t(z)}{\partial t} \right| < n^{\frac{2}{3}-\delta}$$

for all  $x \in [E_1, E_2]$ . From the definitions in (6.2) we have

$$\frac{\partial \text{Tr} \widehat{G}_t}{\partial t} = -\text{Tr} \left( \widehat{G}_t \left( (HU' + UH') + 2t \left( HH' - \frac{1}{3} \right) \right) \widehat{G}_t \right).$$

Hence for (6.4) it suffices to show the following estimates hold for all  $x \in [E_1, E_2]$ :

$$(6.5) \quad \begin{aligned} & |\text{Tr}(HU' \widehat{G}_t^2)| < n^{\frac{2}{3}-\delta}, \quad |(\text{Tr} UH' \widehat{G}_t^2)| < n^{\frac{2}{3}-\delta}, \\ & \left| \text{Tr} \left( \left( HH' - \frac{1}{3} \right) \widehat{G}_t^2 \right) \right| < n^{\frac{2}{3}-\delta}. \end{aligned}$$

We start with the first estimate in (6.5). The other two can be derived similarly. Let

$$(6.6) \quad \mathcal{P} := \text{Tr}(HU' \widehat{G}_t^2), \quad \mathbf{m}^{(k,\ell)} := \mathcal{P}^k \overline{\mathcal{P}}^\ell.$$

Our aim is to establish the following recursive moment estimate: for any fixed integer  $k > 0$

$$(6.7) \quad \mathbb{E}(\mathbf{m}^{(k,k)}) = \mathbb{E}(\mathbf{c}_1 \mathbf{m}^{(k-1,k)}) + \mathbb{E}(\mathbf{c}_2 \mathbf{m}^{(k-2,k)}) + \mathbb{E}(\mathbf{c}_3 \mathbf{m}^{(k-1,k-1)})$$

for some random quantities  $\mathbf{c}_i, i = 1, 2, 3$  which satisfy

$$(6.8) \quad |\mathbf{c}_1| < n^{\frac{2}{3}-\delta}, \quad |\mathbf{c}_2| < n^{\frac{4}{3}-2\delta}, \quad |\mathbf{c}_3| < n^{\frac{4}{3}-2\delta},$$

$$(6.9) \quad \mathbb{E}|\mathbf{c}_1|^{2k} < n^{2k(\frac{2}{3}-\delta)}, \quad \mathbb{E}|\mathbf{c}_2|^k < n^{2k(\frac{2}{3}-\delta)}, \quad \mathbb{E}|\mathbf{c}_3|^k < n^{2k(\frac{2}{3}-\delta)}.$$

Assuming (6.7), by Young’s inequality we have, for any given small  $\varepsilon$ ,

$$\mathbb{E}(\mathbf{m}^{(k,k)}) \leq 3 \frac{1}{2k} n^{2k\varepsilon} n^{2k(\frac{2}{3}-\delta)} + 3 \frac{2k-1}{2k} n^{-\frac{2k\varepsilon}{2k-1}} \mathbb{E}(\mathbf{m}^{(k,k)}).$$

Since  $k$  can be any large (but fixed) positive integer, we can conclude the first estimate in (6.5) by applying Markov’s inequality. The above strategy of recursive moment estimate is inspired by a similar idea used in [26].

Hence, what remains is to prove (6.7). In the sequel, for brevity, we only keep tracking the bounds in (6.8). Those in (6.9) will follow easily from (6.8), the deterministic bounds of the entries of  $G$  and  $U$ , together with the Gaussian tail of the entries in  $H$ . To this end, we first use the integration by parts formula for Gaussian random variable

$$(6.10) \quad \begin{aligned} \mathbb{E}(\mathbf{m}^{(k,k)}) &= \mathbb{E}(\text{Tr} HU' \widehat{G}_t^2 \mathbf{m}^{(k-1,k)}) = \sum_{a,(ij)} \mathbb{E}(h_{a,(ij)} (U' \widehat{G}_t^2)_{(ij),a} \mathbf{m}^{(k-1,k)}) \\ &= \frac{1}{3M} \sum_{a,(ij)} \mathbb{E} \left( \frac{\partial (U' \widehat{G}_t^2)_{(ij),a}}{\partial h_{a,(ij)}} \mathbf{m}^{(k-1,k)} \right) \\ &\quad + \frac{k-1}{3M} \sum_{a,(ij)} \mathbb{E} \left( (U' \widehat{G}_t^2)_{(ij),a} \frac{\partial \mathcal{P}}{\partial h_{a,(ij)}} \mathbf{m}^{(k-2,k)} \right) \\ &\quad + \frac{k}{3M} \sum_{a,(ij)} \mathbb{E} \left( (U' \widehat{G}_t^2)_{(ij),a} \frac{\partial \overline{\mathcal{P}}}{\partial h_{a,(ij)}} \mathbf{m}^{(k-1,k-1)} \right). \end{aligned}$$

Here we use the notation  $\sum_{a,(ij)}$  to represent the sum over  $a \in \llbracket 1, p \rrbracket$ ,  $1 \leq i < j \leq n$ . Hence, to establish (6.7), it suffices to show

$$\begin{aligned}
 & \frac{1}{M} \sum_{a,(ij)} \frac{\partial(U' \widehat{G}_t^2)_{(ij),a}}{\partial h_{a,(ij)}} = O_{\prec}(n^{\frac{2}{3}-\delta}), \\
 (6.11) \quad & \frac{1}{M} \sum_{a,(ij)} (U' \widehat{G}_t^2)_{(ij),a} \frac{\partial \mathcal{P}}{\partial h_{a,(ij)}} = O_{\prec}(n^{\frac{4}{3}-2\delta}), \\
 & \frac{1}{M} \sum_{a,(ij)} (U' \widehat{G}_t^2)_{(ij),a} \frac{\partial \overline{\mathcal{P}}}{\partial h_{a,(ij)}} = O_{\prec}(n^{\frac{4}{3}-2\delta}).
 \end{aligned}$$

The proofs of the last two estimates are similar. Hence, we only show the details of the proofs for the first two estimates above. Set  $\widehat{\Theta}_t := U + tH$ . It is easy to obtain from (6.2) that

$$\frac{\partial \widehat{G}_t}{\partial h_{a,(ij)}} = -t \widehat{G}_t (E_{a,(ij)} \widehat{\Theta}'_t + \widehat{\Theta}_t (E_{a,(ij)})') \widehat{G}_t,$$

where we use the notation  $E_{a,(ij)}$  to denote the  $p \times M$  matrix whose  $(a, (ij))$ th entry is 1 and all the other entries are 0. Then, it is easy to check

$$\begin{aligned}
 \frac{\partial(U' \widehat{G}_t^2)_{(ij),a}}{\partial h_{a,(ij)}} &= -t(U' \widehat{G}_t)_{(ij),a} (\widehat{\Theta}'_t \widehat{G}_t^2)_{(ij),a} - t(U' \widehat{G}_t \widehat{\Theta}_t)_{(ij)(ij)} (\widehat{G}_t^2)_{aa} \\
 &\quad - t(U' \widehat{G}_t^2)_{(ij),a} (\widehat{\Theta}'_t \widehat{G}_t)_{(ij),a} - t(U' \widehat{G}_t^2 \widehat{\Theta}_t)_{(ij)(ij)} (\widehat{G}_t)_{aa}
 \end{aligned}$$

and

$$\begin{aligned}
 \frac{\partial \mathcal{P}}{\partial h_{a,(ij)}} &= (U' \widehat{G}_t^2)_{(ij),a} - t(\widehat{\Theta}'_t \widehat{G}_t H U' \widehat{G}_t^2)_{(ij),a} - t(\widehat{G}_t H U' \widehat{G}_t^2 \widehat{\Theta}_t)_{a,(ij)} \\
 &\quad - t(\widehat{\Theta}'_t \widehat{G}_t^2 H U' \widehat{G}_t)_{(ij),a} - t(\widehat{G}_t^2 H U' \widehat{G}_t \widehat{\Theta}_t)_{a,(ij)}.
 \end{aligned}$$

Consequently, we have

$$\begin{aligned}
 (6.12) \quad \frac{1}{M} \sum_{a,(ij)} \frac{\partial(U' \widehat{G}_t^2)_{(ij),a}}{\partial h_{a,(ij)}} &= -\frac{t}{M} \text{Tr} \widehat{G}_t^2 \widehat{\Theta}_t U' \widehat{G}_t - \frac{t}{M} \text{Tr} \widehat{\Theta}_t U' \widehat{G}_t \text{Tr} \widehat{G}_t^2 \\
 &\quad - \frac{t}{M} \text{Tr} \widehat{G}_t \widehat{\Theta}_t U' \widehat{G}_t^2 - \frac{t}{M} \text{Tr} \widehat{\Theta}_t U' \widehat{G}_t^2 \text{Tr} \widehat{G}_t
 \end{aligned}$$

and

$$\begin{aligned}
 & \frac{1}{M} \sum_{a,(ij)} (U' \widehat{G}_t^2)_{(ij),a} \frac{\partial \mathcal{P}}{\partial h_{a,(ij)}} \\
 &= \frac{1}{M} \text{Tr} \widehat{G}_t^2 U U' \widehat{G}_t^2 - \frac{t}{M} \text{Tr} \widehat{G}_t^2 U \widehat{\Theta}'_t \widehat{G}_t H U' \widehat{G}_t^2
 \end{aligned}$$

$$\begin{aligned}
 & - \frac{t}{M} \operatorname{Tr} \widehat{G}_t H U' \widehat{G}_t^2 \widehat{\Theta}_t U' \widehat{G}_t^2 \\
 (6.13) \quad & - \frac{t}{M} \operatorname{Tr} \widehat{G}_t^2 U \widehat{\Theta}_t' \widehat{G}_t^2 H U' \widehat{G}_t - \frac{t}{M} \operatorname{Tr} \widehat{G}_t^2 H U' \widehat{G}_t \widehat{\Theta}_t U' \widehat{G}_t^2.
 \end{aligned}$$

Now we claim that

$$(6.14) \quad \|HU'\| \prec n^{-\frac{1}{2}}, \quad \|UU'\| \prec 1.$$

To see the first estimate, we first notice that

$$(6.15) \quad \|HU'UH'\| = \|HT'V'V.TH'\| \prec \frac{1}{n} \|HT'TH'\|,$$

where we use the notation  $V_i$  to represent the  $p \times n$  matrix with  $v_i$  as its  $i$ th row. In the last step, we use the fact that  $V'V$  is a sample covariance matrix with entries (in  $V_i$ ) of order  $\frac{1}{\sqrt{M}} \sim \frac{1}{n}$ , which implies that  $\|V'V\| \prec \frac{1}{n}$  (cf. Proposition S0.6).

Further, observe that  $T'T$  is a rank  $n$  matrix with  $\|T'T\| = \frac{1}{3} \|\Gamma\| = O(n)$ . Writing the spectral decomposition as  $T'T := O_T' \Lambda_T O_T$ , we have the fact that

$$(6.16) \quad \|HT'TH'\| \prec n \|HO_T'(I_n \oplus 0)O_TH'\| \stackrel{d}{=} n \|\mathcal{H}\mathcal{H}'\|,$$

where  $\mathcal{H}$  is a  $p \times n$  matrix with i.i.d.  $N(0, \frac{1}{M})$  entries. Then the first estimate in (6.14) follows simply from the fact that  $\|\mathcal{H}\mathcal{H}'\| \prec \frac{1}{n}$ , (6.16) and (6.15). The second estimate in (6.14) is easy to see from the fact that  $\|U'U\| = \|T'V'V.T\| \prec \frac{1}{n} \|T'T\| \prec 1$ . Then, using (6.14) to (6.12), we have

$$\begin{aligned}
 \left| \frac{1}{M} \sum_{a,(ij)} \frac{\partial (U' \widehat{G}_t^2)_{(ij),a}}{\partial h_{a,(ij)}} \right| & \prec \frac{1}{M} \operatorname{Tr} |\widehat{G}_t|^3 + \frac{1}{M} \operatorname{Tr} |\widehat{G}_t|^2 \operatorname{Tr} |\widehat{G}_t| \\
 & \leq \frac{1}{M\eta^2} \operatorname{Im} \operatorname{Tr} \widehat{G}_t + \frac{1}{M\eta} \operatorname{Im} \operatorname{Tr} \widehat{G}_t \operatorname{Tr} |\widehat{G}_t| \prec n^{\frac{1}{3}+\varepsilon},
 \end{aligned}$$

where in the last step we use the local laws Lemma 6.2 and Lemma S0.5.

Similarly, using (6.13) and (6.14), we have

$$\begin{aligned}
 \left| \frac{1}{M} \sum_{a,(ij)} (U' \widehat{G}_t^2)_{(ij),a} \frac{\partial \mathcal{P}}{\partial h_{a,(ij)}} \right| & \prec \frac{1}{M} \operatorname{Tr} |\widehat{G}_t|^4 + \frac{1}{M\sqrt{n}} \operatorname{Tr} |\widehat{G}_t|^5 \\
 (6.17) \quad & \leq \frac{1}{M\eta^3} \operatorname{Im} \operatorname{Tr} \widehat{G}_t + \frac{1}{M\sqrt{n}\eta^4} \operatorname{Im} \operatorname{Tr} \widehat{G}_t \prec n^{\frac{5}{6}+\varepsilon},
 \end{aligned}$$

where again in the last step we use the local laws Lemma 6.2 and Lemma S0.5. Hence, we conclude the proof of the first two estimates in (6.11). The last one can be proved similarly to the second one, we thus omit the details. Therefore, we get (6.7). Then, by Young's inequality, we can get the first estimate in (6.5). The second estimate in (6.5) can be proved analogously and thus we omit the details. For the last estimate in (6.5), we can also use the same strategy, and the details of its proof is stated in the Supplementary Material [7]. Therefore, we completed the proof of Proposition 6.1.  $\square$

**7. Edge universality for  $K$ .** With Propositions 5.1 and 6.1, we can now prove Theorem 1.2 and Corollary 1.4.

PROOF OF THEOREM 1.2. Using Propositions 5.1 and 6.1 we obtain

$$(7.1) \quad \left| \mathbb{E}F \left( n \int_{E_1}^{E_2} \operatorname{Im} m(x + \lambda_{+,c_n} + i\eta) dx \right) - \mathbb{E}F \left( n \int_{E_1}^{E_2} \operatorname{Im} \tilde{m}(x + \lambda_{+,c_n} + i\eta) dx \right) \right| \leq n^{-\delta},$$

where  $F$ ,  $E_1$ ,  $E_2$  and  $\eta$  satisfy the assumptions in Proposition 5.1. Similar to the proof of Theorem 1.1 in [31], one can show by using (7.1) and the local laws that

$$(7.2) \quad \begin{aligned} & \mathbb{P} \left( n^{\frac{2}{3}} (\lambda_1(K) - \lambda_{+,c_n}) \leq s - n^{-\varepsilon} \right) - n^{-\delta} \\ & \leq \mathbb{P} \left( n^{\frac{2}{3}} (\lambda_1(\tilde{K}) - \lambda_{+,c_n}) \leq s \right) \\ & \leq \mathbb{P} \left( n^{\frac{2}{3}} (\lambda_1(K) - \lambda_{+,c_n}) \leq s + n^{-\varepsilon} \right) + n^{-\delta}. \end{aligned}$$

Further, we observe that  $UU' = V.TT'V$ . In addition, we notice that  $TT' = nI_n - \mathbf{1}\mathbf{1}'$ . Denoting by  $\mathcal{V} := \sqrt{\frac{3}{2}(n-1)}V$ , and  $\Sigma = I_n - \frac{1}{n}\mathbf{1}\mathbf{1}'$ , we can write

$$(7.3) \quad \tilde{K} = UU' + \frac{1}{3}I_p = \frac{2n}{3(n-1)}\mathcal{V}\Sigma\mathcal{V}' + \frac{1}{3}I_p.$$

It is known from Theorem 2.7 of [9] that the largest eigenvalues of  $\mathcal{V}\Sigma\mathcal{V}'$  differ from the corresponding ones of  $\mathcal{V}\mathcal{V}'$  only by  $O_{\prec}(\frac{1}{n})$ . This together with Theorem 1.1 in [31] leads to

$$(7.4) \quad \begin{aligned} & \mathbb{P} \left( \frac{3}{2}n^{\frac{2}{3}} (\lambda_1(\tilde{K}) - \lambda_{+,c_n}) \leq s - n^{-\varepsilon} \right) - n^{-\delta} \\ & \leq \mathbb{P} \left( n^{\frac{2}{3}} (\lambda_1(Q) - d_{+,c_n}) \leq s \right) \\ & \leq \mathbb{P} \left( \frac{3}{2}n^{\frac{2}{3}} (\lambda_1(\tilde{K}) - \lambda_{+,c_n}) \leq s \right) + n^{-\delta}. \end{aligned}$$

Combining (7.2) and (7.4), we obtain (1.7). This concludes the proof.  $\square$

PROOF OF COROLLARY 1.4. The conclusion follows directly from Theorem 1.2 and the Tracy–Widom limit for  $\lambda_1(Q)$  (cf. [21]).  $\square$

**Acknowledgments.** The author would like to thank Jiang Hu and Wang Zhou for helpful discussion.

SUPPLEMENTARY MATERIAL

**Supplement to “Tracy–Widom limit for Kendall’s tau”** (DOI: [10.1214/18-AOS1786SUPP](https://doi.org/10.1214/18-AOS1786SUPP); .pdf). The supplement includes the proofs of some technical lemmas and some additional simulation studies.

## REFERENCES

- [1] BAI, Z. and ZHOU, W. (2008). Large sample covariance matrices without independence structures in columns. *Statist. Sinica* **18** 425–442. [MR2411613](#)
- [2] BAIK, J., BEN AROUS, G. and PÉCHÉ, S. (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.* **33** 1643–1697. [MR2165575](#)
- [3] BANDEIRA, A. S., LODHIA, A. and RIGOLLET, P. (2017). Marčenko–Pastur law for Kendall’s tau. *Electron. Commun. Probab.* **22** Paper no. 32. [MR3663103](#)
- [4] BAO, Z., LIN, L.-C., PAN, G. and ZHOU, W. (2015). Spectral statistics of large dimensional Spearman’s rank correlation matrix and its application. *Ann. Statist.* **43** 2588–2623. [MR3405605](#)
- [5] BAO, Z., PAN, G. and ZHOU, W. (2012). Tracy–Widom law for the extreme eigenvalues of sample correlation matrices. *Electron. J. Probab.* **17** Paper no. 88. [MR2988403](#)
- [6] BAO, Z., PAN, G. and ZHOU, W. (2015). Universality for the largest eigenvalue of sample covariance matrices with general population. *Ann. Statist.* **43** 382–421. [MR3311864](#)
- [7] BAO, Z. (2019). Supplement to “Tracy–Widom limit for Kendall’s tau.” DOI:10.1214/18-AOS1786SUPP.
- [8] BAO, Z. G. (2018). Tracy–Widom limit for Spearman’s rho. Preprint.
- [9] BLOEMENDAL, A., KNOWLES, A., YAU, H.-T. and YIN, J. (2016). On the principal components of sample covariance matrices. *Probab. Theory Related Fields* **164** 459–552. [MR3449395](#)
- [10] DING, X. and YANG, F. (2018). A necessary and sufficient condition for edge universality at the largest singular values of covariance matrices. *Ann. Appl. Probab.* **28** 1679–1738. [MR3809475](#)
- [11] EL KAROUI, N. (2006). A rate of convergence result for the largest eigenvalue of complex white Wishart matrices. *Ann. Probab.* **34** 2077–2117. [MR2294977](#)
- [12] EL KAROUI, N. (2007). Tracy–Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices. *Ann. Probab.* **35** 663–714. [MR2308592](#)
- [13] ERDŐS, L., KNOWLES, A. and YAU, H.-T. (2013). Averaging fluctuations in resolvents of random band matrices. *Ann. Henri Poincaré* **14** 1837–1926. [MR3119922](#)
- [14] ERDŐS, L., YAU, H.-T. and YIN, J. (2012). Bulk universality for generalized Wigner matrices. *Probab. Theory Related Fields* **154** 341–407. [MR2981427](#)
- [15] ERDŐS, L., YAU, H.-T. and YIN, J. (2012). Rigidity of eigenvalues of generalized Wigner matrices. *Adv. Math.* **229** 1435–1515. [MR2871147](#)
- [16] FAN, Z. and JOHNSTONE, I. (2017). Tracy–Widom at each edge of real covariance estimators. Available at [arXiv:1707.02352](#).
- [17] GAO, J., HAN, X., PAN, G. and YANG, Y. (2017). High dimensional correlation matrices: The central limit theorem and its applications. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 677–693. [MR3641402](#)
- [18] HOEFFDING, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Stat.* **19** 293–325. [MR0026294](#)
- [19] JIANG, T. (2004). The asymptotic distributions of the largest entries of sample correlation matrices. *Ann. Appl. Probab.* **14** 865–880. [MR2052906](#)
- [20] JOHANSSON, K. (2000). Shape fluctuations and random matrices. *Comm. Math. Phys.* **209** 437–476. [MR1737991](#)
- [21] JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. [MR1863961](#)
- [22] JOHNSTONE, I. M. and MA, Z. (2012). Fast approach to the Tracy–Widom law at the edge of GOE and GUE. *Ann. Appl. Probab.* **22** 1962–1988. [MR3025686](#)
- [23] KNOWLES, A. and YIN, J. (2017). Anisotropic local laws for random matrices. *Probab. Theory Related Fields* **169** 257–352. [MR3704770](#)

- [24] LEE, J. O. and SCHNELLI, K. (2015). Edge universality for deformed Wigner matrices. *Rev. Math. Phys.* **27** 1550018. [MR3405746](#)
- [25] LEE, J. O. and SCHNELLI, K. (2016). Tracy–Widom distribution for the largest eigenvalue of real sample covariance matrices with general population. *Ann. Appl. Probab.* **26** 3786–3839. [MR3582818](#)
- [26] LEE, J. O. and SCHNELLI, K. (2018). Local law and Tracy–Widom limit for sparse random matrices. *Probab. Theory Related Fields* **171** 543–616. [MR3800840](#)
- [27] MA, Z. (2012). Accuracy of the Tracy–Widom limits for the extreme eigenvalues in white Wishart matrices. *Bernoulli* **18** 322–359. [MR2888709](#)
- [28] MARCHENKO, V. A. and PASTUR, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Math. USSR, Sb.* **1** 457.
- [29] ONATSKI, A. (2008). The Tracy–Widom limit for the largest eigenvalues of singular complex Wishart matrices. *Ann. Appl. Probab.* **18** 470–490. [MR2398763](#)
- [30] PILLAI, N. S. and YIN, J. (2012). Edge universality of correlation matrices. *Ann. Statist.* **40** 1737–1763. [MR3015042](#)
- [31] PILLAI, N. S. and YIN, J. (2014). Universality of covariance matrices. *Ann. Appl. Probab.* **24** 935–1001. [MR3199978](#)
- [32] WANG, K. (2012). Random covariance matrices: Universality of local statistics of eigenvalues up to the edge. *Random Matrices Theory Appl.* **1** 1150005. [MR2930383](#)
- [33] ZHOU, W. (2007). Asymptotic distribution of the largest off-diagonal entry of correlation matrices. *Trans. Amer. Math. Soc.* **359** 5345–5363. [MR2327033](#)

DEPARTMENT OF MATHEMATICS  
HONG KONG UNIVERSITY OF SCIENCE  
AND TECHNOLOGY  
HONG KONG  
E-MAIL: [mazgbao@ust.hk](mailto:mazgbao@ust.hk)